Titre: Title:	Synthèse de vues par appariement épars et triangulation
Auteur: Author:	Jean-Sébastien Perrier
Date:	2002
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Perrier, J.-S. (2002). Synthèse de vues par appariement épars et triangulation [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/6989/

UNIVERSITÉ DE MONTRÉAL

SYNTHÈSE DE VUES PAR APPARIEMENT ÉPARS ET TRIANGULATION

JEAN-SÉBASTIEN PERRIER
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE ET INFORMATIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
MAI 2002

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

SYNTHÈSE DE VUES PAR APPARIEMENT ÉPARS ET TRIANGULATION

présenté par: PERRIER Jean-Sébastien,

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de:

Mme CHERIET Farida, Ph.D., présidente

M. COHEN Paul, Ph.D., directeur de recherche

M. AGAM Gady, Ph.D., codirecteur de recherche

M. BRAULT Jean-Jules, Ph.D., membre

Remerciements

Je voudrais remercier M. Paul Cohen de m'avoir accueilli dans le Groupe de Recherche en Perception et Robotique et de m'avoir encadré tout au long de mes travaux de recherche. Je remercie également M. Gady Agam, qui m'a supervisé dans mon travail et qui a contribué grandement à l'élaboration et à la continuité de ce projet. Je remercie aussi les membres du jury pour le temps qu'ils auront consacré à l'évaluation de ce travail.

De plus, je suis reconnaissant envers les Fonds FCAR pour la bourse de recherche qui m'a été attribuée au début de mes études, et envers la chaire CRSNG-Noranda pour l'appui financier qui m'a été accordé dans le cadre du projet SMART.

J'aimerais exprimer ma reconnaissance envers tous les membres (passés et présents) du GRPR pour leur aide, leurs encouragements et leur amitié. En particulier, je voudrais remercier Guy Michaud, pour sa collaboration au début de mes travaux de recherche, ainsi que Claude Duchesne pour avoir entretenu mon équilibre mental et physique pendant nos innombrables parties de badminton.

Résumé

La synthèse de vues basée sur des images est encore un sujet de recherche récent, mais dont les applications sont déjà très nombreuses et ne cessent de se multiplier. La possibilité de créer de nouvelles vues d'une scène à partir d'images déjà existantes permet la génération d'environnements synthétiques et de réalité augmentée. Les différentes approches pour la synthèse de vues reposent, entre autre, sur l'établissement de mise en correspondance entre les images utilisées. Cet appariement peut être de deux types, soit dense ou bien épars. L'appariement dense consiste à établir la correspondance entre tous les pixels des images alors que l'appariement épars ne va établir de correspondance que là où il y a des points d'intérêt (c'est-à-dire des arêtes ou des coins). Dans les deux cas, certaines parties des images demeurent inévitablement non appariées puisqu'il n'y a pas toujours d'information partout dans les images pour supporter un appariement dense.

La triangulation des points d'appariement épars se présente comme une solution possible pour gérer ces zones non appariées. La triangulation consiste à créer des triangles connectant les différents points d'appariement. On fait ainsi la supposition que les triangles représentent des surfaces planes dans la scène observée. Pour être utilisable, une telle méthode de triangulation doit respecter la géométrie de la scène qu'elle tente de modéliser. On entend par là que les triangles doivent cor-

respondent à des surfaces réelles dans la scène. Ce mémoire propose une nouvelle approche pour la triangulation physiquement valide des points d'appariement épars. L'approche proposée est basée sur la maximisation d'un critère de validité physique qui est mesuré sur les régions texturées des images. La triangulation ainsi produite est telle que chaque triangle correspond approximativement à une surface plane dans la scène. Étant donné une triangulation initiale arbitraire, l'approche suggérée utilise un opérateur de basculement d'arêtes afin de modifier les connections entre les points et ainsi se déplacer dans l'espace des configurations des triangles. Toutefois, certains points d'appariement manquant à des endroits spécifiques peuvent empêcher la triangulation de respecter la géométrie de la scène. Un algorithme de raffinement de la triangulation est donc proposé. Ce dernier ajoute des points d'appariement à l'intérieur des triangles qui ne sont pas physiquement valides. La validité des points ajoutés est évaluée avec la mesure de planarité des triangles. Ainsi, la région de support de corrélation utilisée est beaucoup plus grande que celle utilisée habituellement pour la mise en correspondance éparse.

Pour démontrer la validité de la méthode proposée, un ensemble de tests est présenté. Ces tests ont pour but l'évaluation de la qualité de la triangulation produite dans différentes conditions. Ces conditions comprennent des rotations en 3D autour d'un axe vertical qui modifient la profondeur de la surface observée, et également l'utilisation de différents types de texture et de géométrie de la scène. De plus, les résultats de la triangulation sont utilisés avec différentes méthodes de synthèse de vues pour démontrer la qualité des images générées. Spécifiquement, les méthodes d'interpolation et d'extrapolation de vues sont utilisées de façon très satisfaisante avec les ensembles de triangles produits par la méthode de triangulation proposée.

Abstract

Image-based view synthesis is an emerging research topic, with numerous applications. The ability to synthesize new views from existing images, enables the generation of enhanced and synthetic environments. Approaches for view synthesis relay upon dense or sparse matching of the views. In both cases, some parts of the images are inevitably unmatched. Triangulation of sparsely matched points present a possible solution for the handling of those unmatched regions. However, such triangulation should respect the underlying geometry of the scene. In this work, a novel approach is proposed for physically valid triangulation of sparsely matched points. The proposed approach is based on the maximization of a physical validity criterion which is supported by textured regions in the images. The produced triangulation is such that each triangle corresponds approximately to a planar surface in the scene. Given an arbitrary initial triangulation, the proposed approach refines it by flipping the edges of triangles. Furthermore, since missing matched points may preclude the correct triangulation of the scene, an additional stage handles the addition of matched points inside low score triangles. Inherent to this approach, the support region which is used for the evaluation of the correctness of the added matches is normally much larger than the one used in a local match evaluation. A set of tests where conducted to evaluate the quality of the resulting triangulation under different conditions, such as in depth

rotations of the observed scene and also, different kind of scenes and textures. Since the principal application of this method is image-based view synthesis, the results of the triangulation where demonstrated using different two kinds of methods, view interpolation and view extrapolation. The results of the method are visually satisfying for both view synthesis methods.

Table des matières

Liste des figures

Chapitre 1

Introduction

1.1 Synthèse de vues basée sur des images

Depuis les dernières années, la synthèse de vues basée sur des images a fait l'objet de recherches intensives pour plusieurs applications qui s'étendent du rendu de qualité photographique de scènes virtuelles complexes ou de vraies scènes à la compression efficace de séquences vidéo. De façon plus spécifique, le présent travail vise à utiliser la synthèse de vue afin de générer des images pour un télé-opérateur de machineries robotisées. Dans un environnement hostile, comme les mines souterraines, par mesure de sécurité, l'opérateur ne peut pas voir le robot directement et doit utiliser des caméras posés sur ce dernier. Toutefois, un signal vidéo est très difficile à transmettre dans de telles conditions à cause de la faible bande passante du système de transmission. Seules quelques images peuvent être transmises de façon intermittente. De plus, les caméras ne sont pas toujours placées d'une façon idéale pour que l'opérateur puisse spécifier la tâche à accomplir. Le présent travail vise donc à utiliser la synthèse de vues basée sur des images pour reconstituer l'environnement du robot et permettre

à l'opérateur de placer la caméra virtuelle selon un point de vue qui facilitera la spécification de la tâche.

Par définition, la synthèse de vues basée sur des images est un processus qui prend des images d'une scène en entrée et produit de nouvelles images de cette même scène depuis différents points de vue en sortie. Ce processus est très complexe en soit et peut être décomposé en sous-processus. La figure 1.1 illustre une telle décomposition du processus de rendu basé sur des images. Les images fournies en entrée sont envoyées à un processus de modélisation basé sur des images. Celui-ci va extraire l'information des images fournies et la mettre sous la forme requise pour le processus de rendu. La forme et la complexité du modèle de la scène ainsi généré dépendent de la méthode qui est utilisée. Par exemple, le modèle de la scène peut être constitué d'un ensemble d'images reliées les unes aux autres par des groupes de points d'appariement ainsi qu'éventuellement, des matrices de transformation. Il peut également être une reconstruction 3D partielle ou complète de la scène. Le processus de modélisation utilise des techniques d'appariement et, parfois, de reconnaissance de forme. Ce processus peut être complètement automatique ou bien semi-automatique avec des interventions de la part d'un usager.

Le modèle de la scène sert d'intermédiaire entre le processus de modélisation et le processus de rendu. On peut ainsi obtenir une différence de fréquence d'exécution entre les deux processus sans problème. En effet, le modèle de la scène peut être utilisé pour plusieurs itérations du processus de rendu alors que le processus de modélisation complète une seule itération. Le processus d'affichage peut être très rapide si, par exemple, le modèle de la scène est constitué d'un ensemble de polygones 3D. Dans ce cas, le processus d'affichage peut utiliser les cartes graphiques accélératrices pour afficher la scène à des vitesses suffisantes (c'est-à-dire, de 30 à 60 images par secondes)

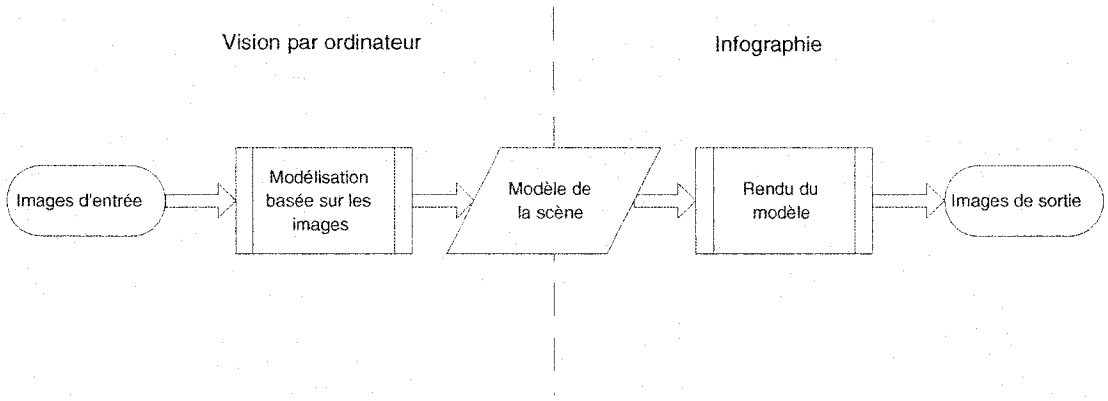


Figure 1.1 : Diagramme de flux de données du processus d'affichage basé sur des images.

pour en permettre un usage interactif.

Toutes les méthodes de synthèse de vues basée sur des images sont structurées comme il est illustré dans la figure 1.1. Le processus de modélisation basé sur les images utilise des techniques provenant du domaine de la vision par ordinateur, tel que l'extraction de points d'intérêt, l'appariement, la relaxation, la segmentation, etc. De l'autre côté, le processus d'affichage ou de rendu utilise des techniques provenant du domaine de l'infographie, telles que l'application de texture (*texture mapping*, en anglais), les projections en perspective, la pixelisation (*rasterisation*, en anglais), etc. Donc, la génération de vue basée sur des images est un processus qui couvre deux domaines de la science informatique, et le lien qui unit ces deux champs est le modèle de la scène. Plusieurs techniques peuvent sembler ne pas en utiliser, mais, le simple fait d'avoir des points d'appariements entre deux images peut être considéré comme un modèle de scène implicite. Les données elles-mêmes contenues dans le modèle peuvent prendre plusieurs formes, mais, en fin de compte, il s'agit toujours d'un modèle de la scène.

Une forme de modèle particulièrement utile est le modèle polygonal. Il est com-

posé d'un ensemble de surfaces planes qui produisent une approximation de la scène. Puisque ce modèle est utilisé pour générer de nouvelles images, la précision n'a pas besoin d'être plus grande que la résolution de l'image finale. La seule contrainte est que le modèle soit suffisamment précis, en matière d'approximation de la vraie scène, pour produire des images qui seront semblables, selon les critères de similitude définis pour l'application donnée, à la scène réelle selon le point de vue sélectionné. Dans ce travail, il sera montré qu'une telle représentation peut être obtenue simplement à partir d'un ensemble de points d'appariement entre deux images. L'avantage principal de cette forme de représentation est qu'elle ne nécessite pas de mise en correspondance dense des images. Cette dernière étant difficile à obtenir et le pourcentage d'erreurs par point d'appariement est toujours très élevé. Par contre, un appariement épars peut être obtenu plus efficacement et de façon plus fiable en utilisant des techniques robustes telles que celle décrite dans (Zhang, Deriche, Faugeras et Luong, 1994), où une méthode de relaxation est utilisée en conjonction avec la détermination robuste de la géométrie épipolaire, afin de mettre en correspondance les points d'intérêts extraits des images sources. Un modèle polygonal de la scène peut être construit à partir d'un ensemble de points d'appariement judicieusement sélectionné. Les scènes, particulièrement celles faites par l'homme, peuvent souvent être simplifiées de façon satisfaisante par un ensemble de surfaces planes relativement petites dans l'espace de coordonnées de l'image. Par exemple, un objet dont le relief visible est relativement plat peut être correctement réduit à une simple surface plane dans le modèle de la scène si ce dernier est observé de suffisement loin. Donc, une triangulation correcte d'un ensemble de points épars sélectionné avec soin peut être suffisant pour la synthèse de vues. De plus, la représentation polygonale est plus facile à afficher en utilisant les pipelines graphiques standards communs dans les cartes accélératrices

3D courantes.

1.2 Résumé de la contribution de la thèse

Ce travail présente une nouvelle approche pour obtenir une représentation polygonale physiquement valide d'une scène en utilisant au minimum deux images et un ensemble réduit de points d'appariement épars entre ces dernières. Cet ensemble de points peut être obtenu soit automatiquement ou manuellement, selon les requis de l'appication. La méthode proposée utilise une triangulation de Delaunay des points initiaux comme première approximation de la triangulation désirée. Par la suite, la triangulation est modifiée pour améliorer le critère de validité physique basé sur les images. Cette nouvelle approche est également présentée dans (Perrier, Agam et Cohen, 2000*a*; Perrier, Agam et Cohen, 2000*b*).

La méthode proposée est basée sur la mesure de planarité d'un triangle. Cette mesure évalue la validité physique de chaque triangle individuel dans l'ensemble de triangles. Ceci est fait sans utiliser de calibration d'aucune sorte et n'a pas besoin de mise en correspondance dense. Seul le signal contenu dans les images de chaque triangle correspondant est utilisé. Cette évaluation de planarité est utilisée dans une série d'algorithmes qui transforment et optimisent la triangulation initiale. Le premier de ces algorithmes, décrit dans la section 4.4.1, utilise la mesure de planarité en tant que critère d'optimisation pour modifier les liens entre les points d'appariement disponibles afin d'améliorer la correspondance de la triangulation avec la vraie scène. Cette optimisation est conduite de façon itérative, ce qui permet de produire des résultats affichables à chaque étape. De plus, ce processus peut être utilisé en conjonction avec d'autres critères d'optimisation, tel que la contrainte épipolaire. Le deuxième

algorithme, décrit dans la section 4.4.2, montre comment raffiner la triangulation déjà existante en divisant les triangles par l'ajout de nouveaux points de correspondance. La mesure de planarité est utilisée en conjonction avec une méthode d'extraction de points d'intérêt pour trouver de nouveaux points d'appariement à l'intérieur d'une paire de triangles en correspondance. Ainsi, un triangle sélectionné est brisé en trois triangles, plus petits et plus précis. Un troisième algorithme, présenté dans la section 4.4.3, utilise la mesure de planarité ainsi que d'autres mesures, telles que la forme et la taille des triangles, afin de fusionner ceux-ci dans le but d'éliminer des triangles invalides ou dégénérés ainsi que des points d'appariement incorrects. Finalement, un algorithme complet est proposé dans la section 4.4.4. Il utilise tous les algorithmes décrits précédemment afin de prendre une triangulation de Delaunay, la transformer et l'affiner jusqu'à obtenir une triangulation physiquement valide. La qualité du résultat final ainsi que son niveau d'affinement est contrôlé par les paramètres de l'algorithme. De plus, d'autres contraintes connues peuvent être appliquées pour augmenter la qualité de la triangulation et des points d'appariement produits. La production d'une représentation polygonale physiquement valide d'une scène peut être utilisée directement pour produire de nouvelles vues de cette scène. Il s'ensuit que cette méthode peut être utilisée directement comme une étape de traitement préalable pour beaucoup de méthodes connues de reprojection utilisées en synthèse de vues, incluant la reconstruction 3D. Un avantage de la représentation polygonale est qu'elle peut gérer automatiquement les régions non appariées des images sources. De plus, elle est basée sur un appariement épars qui est plus facile à obtenir qu'un appariement dense. En effet, l'apparient épars peut êter obtenu sans avoir de calibration précise des images, ce qui est pratiquement impossible pour un apparient dense. Seul les points saillants des images sont utilisés pour la mise en correspondance augmentant ainsi la fiabilité

de chaque point apparié. La représentation polygonale est aussi compatible avec les pipelines d'affichage conventionnels et peut donc être utilisée pour de la synthèse de vues en temps réel.

1.3 Organisation de la thèse

Étant donné que ce travail est basé sur la synthèse de vues, une brève revue des principes fondamentaux qui ont été utilisés est présentée dans le chapitre 2. Ces concepts incluent une description de la géométrie épipolaire et son expression mathématique, la matrice fondamentale. Les méthodes pour estimer cette dernière sont également expliquées. Par la suite, une brève description des méthodes de mise en correspondance est présentée, incluant les méthodes d'appariement dense et épars. Finalement, les deux méthodes de projection, l'interpolation et l'extrapolation de vues, utilisées dans ce travail sont présentées.

Le chapitre 3 décrit les travaux connexes. La première partie décrit les méthodes de synthèse de vues à partir d'images non calibrées ou faiblement calibrées. Ces méthodes sont séparées en trois catégories: l'approche mosaïque, l'interpolation de vue et l'extrapolation de vue à partir d'un point de vue arbitraire. La seconde partie du chapitre parle de certaines autres méthodes déjà utilisées pour créer une représentation polygonale physiquement valide d'une scène à partir d'images. Une revue plus approfondie des travaux connexes peut également être trouvée dans (Agam, Michaud, Perrier, Houle et Cohen, 1999*b*; Agam, Michaud, Perrier, Houle et Cohen, 1999*a*).

Le chapitre 4 présente l'approche proposée pour résoudre le problème de triangulation. Premièrement, les principes d'évaluation de planarité et de modification de la triangulation sont expliqués. Ensuite, le processus de rectification des triangles

est détaillé. La rectification consiste à projeter le contenu d'un triangle d'une image source à l'autre. Pour ce processus de rectification de triangle, un bref test est montré pour comprendre l'effet d'une projection de texture affine et perspective. La mesure de planarité est ensuite expliquée. Finalement, les méthodes de modification de la triangulation, basculement des arêtes, division et fusion des triangles, sont présentées.

Le chapitre 5 présente un ensemble de résultats obtenus à partir de différentes expériences faites en utilisant les méthodes décrites dans le chapitre 4. La première partie du chapitre présente le groupe de tests faits pour évaluer l'influence des rotations en 3D sur le comportement de la mesure de planarité. Ensuite, les résultats des méthodes de basculement d'arêtes et d'affinement de la triangulation sont montrés et discutés. Finalement, des résultats de synthèse de vues, obtenus à partir des méthodes décrites dans la section 2.3 sont montrés pour illustrer l'amélioration apportée par l'usage de la triangulation physiquement correcte.

Pour finir, le chapitre 6 conclue ce mémoire avec un résumé de la méthode proposée ainsi que de ses limites. Les travaux futurs possibles sont également présentés.

Chapitre 2

Revue des concepts fondamentaux de la synthèse de vues

Ce chapitre est consacré aux concepts fondamentaux utilisés en synthèse de vues basée sur des images. Le but recherché est de produire de nouvelles images d'une scène selon des points de vue virtuels. Ces points de vue sont différents de ceux qui ont été utilisés pour obtenir les images réelles de la scène. Les méthodes utilisées pour accomplir la synthèse de vues dépendent des besoins de l'application et du contexte d'utilisation. Par exemple, en télérobotique, la synthèse de nouvelles images peut être utilisée en conjonction avec la réalité augmentée pour permettre à l'opérateur de situer un robot dans son environnement et de lui donner des instructions appropriées. Dans ce cas, l'opérateur peut avoir besoin de déplacer le point de vue virtuel par rapport au point de vue réel afin d'obtenir un meilleur point de vue pour spécifier les paramètres de la tâche. Il faut donc que la méthode de reprojection permette un tel déplacement du point de vue virtuel. On peut également utiliser la synthèse de vues pour des systèmes de vidéo conférence dans lesquels les participants seraient

placés dans un environnement virtuel commun. De tels systèmes existent déjà à l'état expérimental. On peut en trouver un exemple dans (Cooke, Kauff et Schreer, 2002). Dans cet exemple, on utilise la génération de vue intermédiaire en conjonction avec des méthodes d'appariements stéréo pour produire des images rectifiées des participants dans l'environnement virtuel.

La synthèse de nouvelles vues à partir d'images existantes est un processus qui est généralement divisé en trois étapes de traitement des données. La première étape est la calibration des points de vue d'origine. Ceci consiste à établir les caractéristiques géométriques des points de vues qui ont capturé les images sources utilisées. Dans le contexte de la synthèse de vues à partir d'images sans utiliser d'information relative à une application spécifique, on peut obtenir une calibration faible des images. La calibration faible repose sur l'établissement de la géométrie épipolaire des points de vues. Ceci fait l'object de la section 2.1 où l'on présente la géométrie épipolaire et sa représentation mathématique, la matrice fondamentale. On y retrouve également une description de différentes méthodes pour calculer la géométrie épipolaire à partir de points d'appariement.

Une autre étape du processus de synthèse de vues est la mise en correspondance des images sources. Cette étape est souvent accomplie en même temps que la calibration. C'est-à-dire qu'une première mise en correspondance est effectuée de façon grossière à partir de contraintes de calibration très générales. Par exemple, on suppose une disparité maximale d'un quart de la largeur de l'image. Ensuite, une calibration est faite à partir de ces premiers résultats. On obtient ainsi une première itération de résolution du problème couplé calibration - mise en correspondance. On peut ensuite refaire des itérations de raffinement jusqu'à obtenir des résultats plus précis, toujours selon les besoins de l'application. La section 2.2 présente les méthodes de mise en

correspondance utilisées dans ce travail.

Ces deux premières étapes constituent donc l'établissement, ou l'extraction, du modèle de la scène à partir des images sources. Une fois le modèle de la scène établi, il reste finalement à produire de nouvelles images à partir de ce modèle. Ceci est accompli par la reprojection du contenu des images sources vers l'image destination tel qu'observé par le point de vue virtuel désiré. Les méthodes de reprojection utilisées dans ce travail sont décrites dans la section 2.3. La spécification du point de vue virtuel dépend de la méthode utilisée puisque les paramètres contrôlables sont différents selon la méthode choisie.

2.1 Géométrie épipolaire

Dans cette section, la géométrie épipolaire sera expliquée premièrement en décrivant le modèle de caméra *trou d'épingle* qui est utilisé pour modéliser de façon simple une caméra à projection perspective réelle. Ensuite, l'expression mathématique de la géométrie épipolaire, la matrice fondamentale, sera dérivée de l'équation du modèle de caméra *trou d'épingle*. Finalement, les méthodes pour calculer la matrice fondamentale à partir d'un groupe de points d'appariement entre deux projections perspectives seront présentées.

2.1.1 Le modèle de caméra trou d'épingle

Le processus de formation d'une image consiste en la projection de la scène sur un plan. Ce plan est appelé le plan image. Si l'on représente la scène par un nuage de points en 3D, l'image devient la projection de ces points sur le plan image. L'origine du système de coordonnées de la caméra est placée au centre optique (aussi nommé

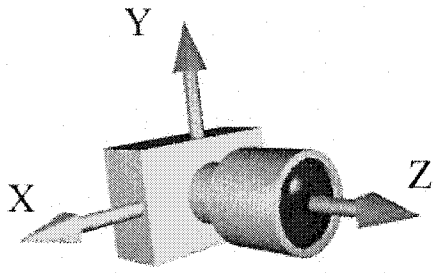


Figure 2.1 : Système de coordonnées de la caméra

centre de projection) de la caméra. L'axe qui sort de l'objectif de la caméra est l'axe Z tel qu'illustré dans la figure 2.1. Le plan image est à une distance f (distance focale) de l'origine du système de coordonnées. L'équation du plan image est donc $Z = f$ dans le système de coordonnées du centre optique de la caméra.

La figure 2.2 illustre la position du plan image par rapport au système de coordonnées de la caméra. La projection d'un point $P = [X, Y, Z]^T$ sur le plan $Z = f$ est un point $p = [x, y, f]^T$ où: $x = \frac{fX}{Z}$ et $y = \frac{fY}{Z}$. En réalité, le centre de projection de la caméra n'est pas toujours aligné avec le centre du plan image et il y a aussi un facteur de mise à l'échelle sur les axes du plan image. Les équations complètes qui donnent les coordonnées d'un point 3D projeté sur le plan image deviennent:

$$x = fS_x\frac{X}{Z} + x_0$$

et

$$y = fS_y\frac{Y}{Z} + y_0$$

où (x_0, y_0) est la projection du centre optique de la caméra dans le plan image. S_x et S_y sont des facteurs de mise à l'échelle pour les axes du plan image. Ces facteurs dépendent principalement des paramètres de fabrication de la caméra réelle utilisée. Afin de manipuler plus aisément ces équations, il est plus pratique de les mettre sous

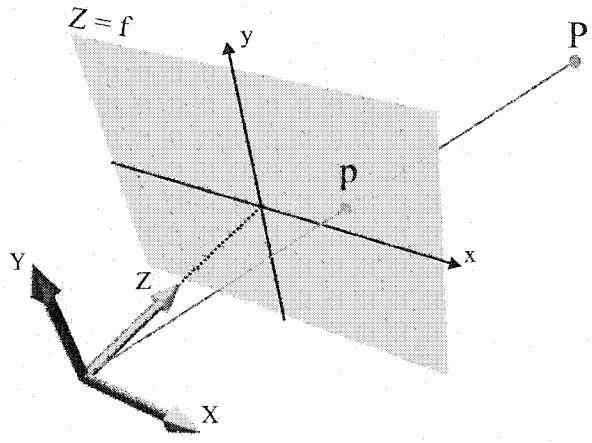


Figure 2.2 : Projection d'un point P dans le plan image de la caméra.

forme matricielle:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} fS_x & 0 & x_0 \\ 0 & fS_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

qui peut être écrit sous une forme plus compacte de la façon suivante:

$$p = \frac{1}{Z} AP \tag{2.1}$$

où

$$A = \begin{bmatrix} fS_x & 0 & x_0 \\ 0 & fS_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

est la matrice des paramètres internes de la caméra. L'équation 2.1 est l'expression mathématique du modèle de caméra trou d'épingle.

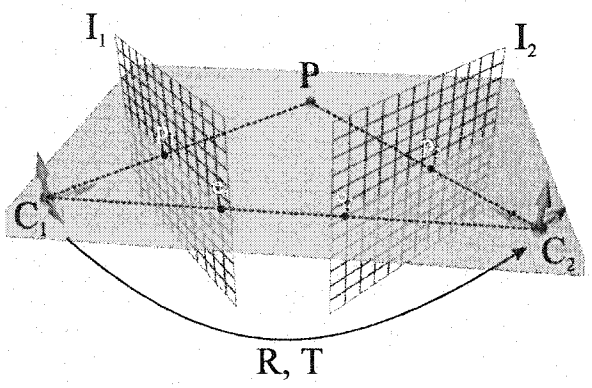


Figure 2.3 : Transformation entre les systèmes de coordonnées des caméras C_1 et C_2. Les points p_1 et p_2 sont les images du point P vu à partir des caméras placées en C_1 et C_2 respectivement.

2.1.2 Matrice fondamentale

Les choses deviennent plus intéressantes lorsqu'il y a deux caméras qui regardent la même scène. Dans un cas général de synthèse de vues, les paramètres externes des caméras sont constitués de la transformation rigide (rotation et translation) qui amène les deux systèmes de coordonnées des caméras l'un sur l'autre. La figure 2.3 montre la transformation rigide entre les deux systèmes de coordonnées des caméras. Une relation existe entre l'image p_1 d'un point 3D P dans le plan image de la caméra 1 et l'image p_2 du même point dans le plan image de la caméra 2. Un plan est formé par les trois points C_1, C_2 et P, qui est appelé le plan épipolaire. Ce plan rencontre les plans images I_1 et I_2 pour former ce que l'on appelle des lignes épipolaires comme on peut le voir dans la figure 2.3. La projection de P dans chaque image est aussi située dans le plan épipolaire et donc, se trouve sur la ligne épipolaire correspondante. De plus, la projection du centre optique d'une caméra dans le plan image de l'autre caméra constitue un point particulier que l'on appelle l'épipole. Toutes les lignes épipolaires d'une image passent forcément par l'épipole. Cette relation entre les deux images d'un point P est appelée la *contrainte épipolaire* et est exprimée mathématiquement

par la matrice fondamentale.

La matrice fondamentale est l'expression algébrique de la contrainte épipolaire. Pour trouver cette expression, il faut tout d'abord exprimer la transformation d'un point P le menant du système de coordonnées de la première caméra vers la seconde:

$$P_2 = RP_1 + T \tag{2.2}$$

où $P_1 = [X_1, Y_1, Z_1]^T$ et $P_2 = [X_2, Y_2, Z_2]^T$ sont les expressions du point P dans les systèmes de coordonnées des caméras C_1 et C_2 respectivement. La rotation entre les systèmes de coordonnées est représentée par une matrice rotation de dimension 3×3:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

et la translation entre les origines des systèmes de coordonnées est exprimée par le vecteur: $T = C_2 - C_1 = [T_X, T_Y, T_Z]^T$. L'application de l'équation 2.1 dans l'équation 2.2, donne:

$$Z_2 A_2^{-1} p_2 = Z_1 R A_1^{-1} p_1 + T$$

En faisant le produit vectoriel par T des deux côtés de l'équation, on obtient:

$$Z_2 T \times A_2^{-1} p_2 = Z_1 T \times R A_1^{-1} p_1 + T \times T$$

Puis, en simplifiant $T \times T = 0$, il s'ensuit que:

$$Z_2 T \times A_2^{-1} p_2 = Z_1 T \times R A_1^{-1} p_1$$

Maintenant, on peut appliquer le produit scalaire par $A_2^{-1}p_2$ des deux côtés pour obtenir:

$$Z_2(T \times A_2^{-1}p_2)A_2^{-1}p_2 = Z_1(T \times RA_1^{-1}p_1)A_2^{-1}p_2$$

Finalement, étant donné que $(T \times A_2^{-1}p_2)A_2^{-1}p_2 = 0$ parce que $T \times (A_2^{-1}p_2)$ est un vecteur orthogonal à $A_2^{-1}p_2$ par définition, on peut éliminer complètement le côté gauche de l'équation. Le facteur Z_1 du côté droite peut également être retiré. On obtient l'expression suivante:

$$(T \times RA_1^{-1}p_1)A_2^{-1}p_2 = 0$$

ou bien, sous une forme plus compacte:

$$p_2^T A_2^{-T} T_S R A_1^{-1} p_1 = 0 \tag{2.3}$$

avec

$$T_S = \text{skew}(T) = \begin{bmatrix} 0 & -T_Z & T_Y \\ T_Z & 0 & -T_X \\ -T_Y & T_X & 0 \end{bmatrix}$$

qui est l'expression matricielle d'un produit vectoriel. D'où on obtient la matrice $F = A_2^{-T} T_S R A_1^{-1}$ qui est une matrice particulière appelée la matrice fondamentale. L'expression matricielle compacte de la contrainte épipolaire est la suivante:

$$p_2^T F p_1 = 0 \tag{2.4}$$

Il est important de prendre note que cette matrice est singulière, c'est-à-dire que $\det(F) = 0$ puisque $\det(T_S) = 0$. En fait, la matrice fondamentale est de rang 2 et,

puisque c'est une matrice projective, elle n'est connue que de façon proportionnelle. Donc, seulement 7 des 9 éléments de la matrice sont linéairement indépendants.

2.1.3 Technique pour estimer la matrice fondamentale

La matrice fondamentale est définie par l'équation 2.4. Pour n'importe quelle paire de points d'appariement $p_1 \leftrightarrow p_2$ dans deux images, la contrainte épipolaire est vérifiée. Étant donné un nombre suffisant de points d'appariement, un système d'équations linéaires peut être utilisé pour calculer la matrice inconnue F. L'écriture au long de l'équation 2.4 donne:

$$\begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0$$

qui peut être écrit sous la forme suivante:

$$uf^T = 0 \tag{2.5}$$

avec

$$u = [x_2x_1, x_2y_1, x_2, y_2x_1, y_2y_1, y_2, x_1, y_1, 1]$$

et

$$f = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}].$$

Étant donné n points de correspondance, On obtient un ensemble d'équations

linéaires que l'on peut exprimer comme:

$$Af = 0 \tag{2.6}$$

où A est une matrice $n \times 9$ qui contient les vecteurs lignes u_i formés par les points de correspondance. Le vecteur solution f est défini seulement de façon proportionnelle. Pour cette raison, et pour prévenir la solution triviale $f = \vec{0}$, un des éléments de f est fixé à 1. Par exemple, on peut dire que $f_{33} = 1$ et résoudre le système d'équations en faisant une minimisation des moindres carrés linéaire. Une autre alternative est d'imposer la contrainte supplémentaire que $\|f\| = f^T f = 1$. Alors, le problème est de trouver un vecteur f qui minimise $\|Af\|$ avec la contrainte $\|f\| = 1$. La solution de ce problème est le vecteur propre unitaire correspondant à la plus petite valeur propre de $A^T A$. Un algorithme approprié pour trouver les valeurs propres est la décomposition en valeur singulière (DVS, ou "Singular Value Decomposition", SVD, en anglais).

La discussion précédente suppose que les données, c'est-à-dire, les points de correspondance, sont de précision parfaite et ne contiennent pas d'erreur. En réalité, à cause des imprécisions des mesures des points d'appariement, la matrice A n'aura pas déficience de rang. Dans ce cas, il ne sera pas possible de trouver une solution non nulle à l'équation $Af = 0$. Pour résoudre ce problème de minimisation, une méthode robuste doit être utilisée. Les méthodes de la section suivante proposent des solutions à ce problème.

2.1.3.1 Algorithme des huit points normalisés

Dans (Hartley, 1997), un algorithme pour résoudre l'équation 2.6 est proposé. Il est appelé l'algorithme des 8 points normalisés. Il est possible de trouver la matrice

fondamentale avec au moins 8 points de correspondance, d'où le nom de l'algorithme. Dans (Hartley, 1997), il est montré que la résolution du système d'équations linéaires souffre d'un mauvais facteur d'échelle numérique. Les éléments linéaires qui résultent de la matrice A^TA varient de 1 à 10^8 ce qui induit des instabilités numériques et amplifie les erreurs dues au bruit et les imprécisions dans les données. Pour résoudre ce problème, il est proposé de normaliser les données avant de résoudre le système d'équations linéaire. Cette normalisation prend la forme d'une transformation des données telle que la condition de la matrice A^TA soit améliorée. Il est possible d'appliquer des transformations T_1 and T_2 sur chaque groupe de points en entrée. Les nouveaux points de correspondance deviennent donc $\hat{p}_1 = T_1 p_1$ et $\hat{p}_2 = T_2 p_2$. Ensuite, la solution du système d'équation sera une matrice $\hat{F}$ qui vérifie la contrainte épipolaire:

$$\hat{p}_2^T \hat{F} \hat{p}_1 = 0$$

qui peut être écrite par rapport aux ensembles de points originaux comme:

$$p_2^T T_2^T \hat{F} T_1 p_1 = 0$$

On peut donc exprimer la matrice fondamentale pour les points originaux de la façon suivante:

$$F = T_2^T \hat{F} T_1. \tag{2.7}$$

La transformation qui améliore la condition du système d'équations est résumée par les étapes suivantes:

1. Les points sont déplacés pour que leur centroïde soit à l'origine de l'image (au lieu du coin supérieur gauche par exemple).

2. Les points sont mis à l'échelle afin que leur distance moyenne avec l'origine de l'image soit égale à $\sqrt{2}$.

Cette transformation est appliquée sur l'ensemble des points de chaque image de façon indépendante. Le système d'équations 2.6 peut ensuite être résolu avec les points d'appariements normalisés. Puis, la matrice fondamentale qui correspond aux points d'appariement originaux est retrouvée grâce à l'équation 2.7.

L'étape finale de l'algorithme proposé dans (Hartley, 1997) est de forcer la contrainte de singularité de la matrice fondamentale. La matrice F est remplacée par une matrice F' qui minimise la norme de Frobenius $\|F - F'\|$ sujette à la condition $\det F' = 0$. On peut obtenir cette matrice en utilisant la décomposition en valeur singulière (DVS). Dans ce cas, $F = UDV^T$ est la DVS de F, où D est une matrice diagonale $D = \text{diag}(r, s, t)$ qui satisfait la condition $r \geq s \geq t$. La matrice singulière $F' = U\text{diag}(r, s, 0)V^T$ est la matrice fondamentale de rang 2 qui est le plus près de la matrice F trouvée précédemment. Cette étape finale assure que les lignes épipolaires de chaque image se croisent au même point, c'est-à-dire à l'épipole de chaque image.

Cet algorithme est très simple à implanter et donne des résultats comparables aux résultats obtenus par les méthodes non linéaires qui sont beaucoup plus difficiles à implanter. Dans le contexte de ce travail, l'algorithme des 8 points est suffisant pour les besoins de calibration faible. Il a donc été utilisé lorsque la contrainte épipolaire devait être retrouvée à partir de points d'appariement.

2.1.3.2 La moindre médiane des carrés

Un second problème survient lorsque l'on essaie de retrouver la matrice fondamentale à partir d'un ensemble de points d'appariement. Il peut s'introduire des erreurs dans les correspondances de points, c'est-à-dire, des points qui sont mis en correspondance

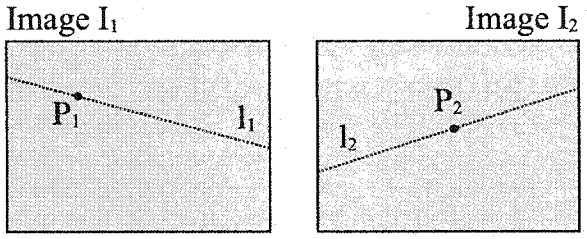


Figure 2.4 : Lignes épipolaires générées par deux points correspondants.

par erreur. Ces points peuvent influencer grandement le résultat de la résolution du système d'équations linéaires. C'est pourquoi, en général, on utilise beaucoup plus que 8 points d'appariement pour calculer la matrice fondamentale. Malgré tout, des points erronés vont rendre toute la solution erronée.

Dans (Zhang et al., 1994), on présente une technique robuste, appelée l'algorithme de la moindre médiane du carré (least median of square, LMedS), qui permet de récupérer la matrice fondamentale tout en excluant les points d'appariement incorrects ou dégénérés. Ceux-ci sont appelés des points excentriques et peuvent affecter sévèrement la précision de la matrice fondamentale s'ils sont utilisés directement dans l'équation 2.6.

Avant de parler de la méthode elle-même, il est nécessaire de définir une formulation pour évaluer les résidus du système linaire. Telle qu'expliqué plus tôt, la contrainte épipolaire implique que chaque point d'une image génère une ligne dans l'autre image, et que le point correspondant dans la seconde image soit sur cette ligne. Dans la figure 2.4, le point p_1 génère une ligne $l_2 = Fp_1$ dans l'image I_2 sur laquelle se trouve le point p_2. De la même façon, le point p_2 génère une ligne $l_1 = F^T p_2$, sur laquelle se trouve le point p_1.

En pratique, les points ne sont pas exactement sur leur ligne épipolaire correspondante. La distance entre le point p_2 et sa ligne épipolaire $l_2 = Fp_1 = [a_2, b_2, c_2]^T$ est

définie comme suit:

$$d(p_2, l_2) = \frac{|p_2^T l_2|}{\sqrt{a_2^2 + b_2^2}}.$$

Toutefois, dans cette définition, les deux images ne jouent pas un rôle symétrique. Il faut donc définir le résidu r_i de l'équation 2.4 comme:

$$r_i = d^2(p_{2i}, l_{2i}) + d^2(p_{1i}, l_{1i}),$$

qui opère simultanément dans les deux images. En utilisant le fait que $p_{2i}^T l_{2i} = p_{1i}^T l_{1i}$, le résidu pour le point i peut être réécrit comme:

$$r_i = \left(\frac{1}{a_{1i}^2 + b_{1i}^2} + \frac{1}{a_{2i}^2 + b_{2i}^2} \right) (p_{2i}^T F p_{1i})^2. \tag{2.8}$$

Avec cette définition de résidu, la méthode LMedS peut être formulée comme la résolution du problème de minimisation non linéaire suivant:

$$\min_i \operatorname{med} r_i.$$

C'est-à-dire que l'estimateur doit produire la plus petite valeur de la médiane des résidus au carré calculée sur la totalité des données. La méthode proposée dans (Zhang et al., 1994) pour résoudre ce problème est de faire un certain nombre d'essais sur un sous-ensemble de données choisi aléatoirement. Étant donnés n points de correspondance, une technique de type Monté Carlo est utilisée pour choisir m sous échantillons aléatoires de 8 points différents. Pour chaque échantillon, numéroté par j, une matrice fondamentale F_j est calculée. Pour chacune de ces matrices, la médiane des résidus au carré, dénotée par M_j, par rapport à l'ensemble complet des données

est calculée:

$$M_j = \underset{i=1,\ldots,n}{\text{med}} \left[d^2(p_{2i}, F_j p_{1i}) + d^2(p_{1i}, F_j^T p_{2i}) \right]$$

Finalement, l'approximation F_j pour laquelle M_j est minimal parmi tous les m sous-échantillons est gardée. Le nombre de sous-échantillons qui doit être essayé dépend de la quantité présumée de points d'appariements corrects. En supposant que l'ensemble des points de correspondance peut contenir une fraction ε de points erronés, la probabilité qu'au moins un des m sous échantillons (qui contient 8 points) ne contienne que des points corrects est donnée par

$$P = 1 - [1 - (1 - \varepsilon)^8]^m.$$

Par exemple, en supposant une fraction $\varepsilon = 0.4$ de points erronés et en nécessitant une probabilité $P = 0.99$ d'avoir au moins un sous-échantillon correct, le nombre de sous-échantillons de 8 points qu'il faudra essayer est de $m = 272$.

D'autres considérations en matière d'efficacité et de robustesse pour cette méthode sont discutées plus à fond dans (Zhang et al., 1994). Il n'est pas nécessaire d'en discuter ici puisque cette méthode a été utilisée pour fin de vérification des résultats seulement.

2.2 Mise en correspondance

Dans cette section, les concepts de base de la mise en correspondance basée sur des images seront couverts. Le but n'est pas de donner une revue complète de tous les travaux de ce domaine, puisque cela dépasserait de loin l'envergure de ce travail. La section 2.2.1 décrit de façon générale les méthodes de mise en correspondance qui

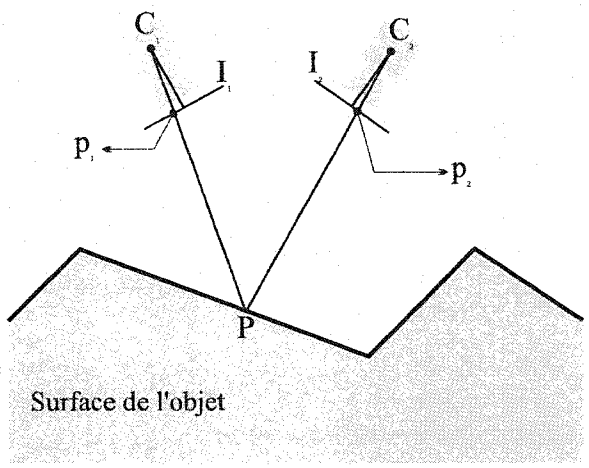


Figure 2.5 : Illustration des projections p_1 et p_2 du point P, dans les images I_1 et I_2 respectivement Les points C_1 et C_2 sont les centres de projection des caméras.

utilisent la corrélation basée sur des images. Ensuite, les deux catégories principales de techniques sont présentées. La première catégorie concerne l'appariement dense, qui est traité dans la section 2.2.2. La deuxième catégorie est composée des méthodes d'appariement épars, qui sont brièvement décrites dans la section 2.2.3.

2.2.1 Corrélation basée sur l'image

Cette section traite de la mise en correspondance en comparant directement les images. Si l'on considère deux images de la même scène, I_1 et I_2, un point p_1 dans I_1 est la projection d'un point P de la scène qui se projette dans I_2 au point p_2 comme l'illustre la figure 2.5. L'appariement basé sur les images est le processus qui établit la correspondance entre la projection d'un même point de la scène dans les deux images. C'est-à-dire, l'identification dans I_2 du pixel (ou du groupe de pixels) qui représente p_2 basé sur le pixel (ou le groupe de pixels) qui représente p_1 dans I_1 et vice versa.

Pour être capable de décider si un point p_2 dans I_2 est la projection du même point que p_1 dans I_1, une mesure de similarité doit être définie. Cette mesure peut être

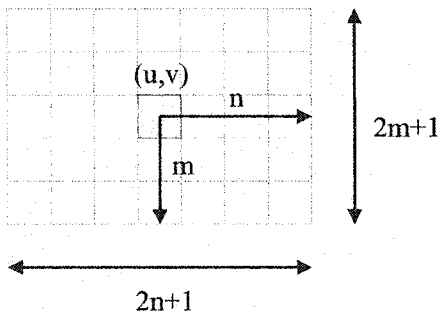


Figure 2.6 : Fenêtre de corrélation de $(2n+1) \times (2m+1)$ pixels autour d'un point $p = [u, v]^T$.

faite en comparant les pixels des images autour de chaque point et en accomplissant une corrélation entre les deux ensembles de pixels. Cette comparaison est valide si les images ont une échelle localement similaire. C'est pourquoi, avant de faire la corrélation, il faut rectifier les images pour rendre leur échelle et leur orientation les plus similaires possibles. La matrice fondamentale (décrite dans la section 2.1.2) peut être utilisée pour établir cette rectification (pour plus de détails sur cette méthode, voir (Seitz et Dyer, 1996*b*)).

Étant donné deux vues rectifiées de la même scène, une corrélation entre deux régions rectangulaires autour de chaque point candidat est accomplie. La brillance d'une image à un point $p_1 = [u_1, v_1]$ est notée comme $I_1(u_1, v_1)$. De la même façon, $I_2(u_2, v_2)$ dénote la brillance de l'image I_2 au point $p_2 = [u_2, v_2]$. La corrélation de la région rectangulaire de $(2n+1) \times (2m+1)$ pixels autour de chaque point, tel que montré dans la figure 2.6, peut être accomplie de plusieurs façons. En pratique, la brillance des mêmes objets varie souvent d'une image à l'autre. Ceci implique que la mesure de corrélation doit compenser pour cette différence.

Une façon de mesurer la différence entre deux vecteurs est de calculer l'angle qui les sépare. Une mesure appelée corrélation normalisée est, en fait, le cosinus de l'angle entre deux vecteurs. Si l'on prend deux vecteurs A et B de dimension N égale, le

cosinus de l'angle entre les deux est:

$$\cos\theta = \frac{A \cdot B}{\|A\|\|B\|}.$$

L'angle θ n'est pas affecté par la longueur des vecteurs. Donc, si les vecteurs A et B sont formés par les brillances de régions rectangulaires des images I_1 et I_2, le résultat de la corrélation normalisée ne sera pas affecté par le facteur d'échelle global d'intensité lumineuse entre les images. En d'autres termes, la corrélation normalisée peut correctement compenser pour des changements d'intensité lumineuse de la forme $I \leftarrow aI$. Dans le cas de régions rectangulaires de taille $(2n+1) \times (2m+1)$, si l'on se base sur l'équation du produit scalaire, la corrélation normalisée (*"Normalized cross-correlation"*, en anglais) peut être écrite comme:

$$\mathrm{NCC}_{mn}(p_1, p_2) = \frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m}[I_1(u_1+i, v_1+j) \cdot I_2(u_2+i, v_2+j)]}{\sqrt{\sum_{i=-n}^{n}\sum_{j=-m}^{m} I_1^2(u_1+i, v_1+j)}\sqrt{\sum_{i=-n}^{n}\sum_{j=-m}^{m} I_2^2(u_2+i, v_2+j)}}. \tag{2.9}$$

Cette mesure de corrélation donne des résultats entre -1 et 1, puisque ce sont les valeurs possibles du cosinus de l'angle entre les vecteurs corrélés.

Il est possible de rendre la corrélation invariante aux changements de luminosité qui consistent en une transformation linéaire, c'est-à-dire des changements de la forme $I \leftarrow aI + b$. Il s'agit de soustraire aux vecteurs la valeur moyenne de leurs éléments avant de faire le produit scalaire. La valeur moyenne d'une région rectangulaire de dimension $m \times n$ pixels dans une image I_k autour d'un point $p_k = [u_k, v_k]$ est donnée par:

$$\mu_k = \frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m} I_k(u_k+i, v_k+j)}{(2n+1)(2m+1)}. \tag{2.10}$$

La mesure qui en résulte est appelée la corrélation normalisée à moyenne nulle (*"zero-*

mean normalized cross-corrélation, ZNCC", en anglais) et est définie par:

$$\mathrm{ZNCC}_{mn}(p_1, p_2) = \frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m}[(I_1(u_1+i, v_1+j) - \mu_1) \cdot (I_2(u_2+i, v_2+j) - \mu_2)]}{V_1(u_1, v_1)V_2(u_2, v_2)} \tag{2.11}$$

où

$$V_k(u_k, v_k) = \sqrt{\sum_{i=-n}^{n}\sum_{j=-m}^{m}(I_k(u_k+i, v_k+j) - \mu_k)^2}.$$

Cette mesure est la plus communément utilisée dans la plupart des méthodes de mise en correspondance utilisant la corrélation basée sur les images.

Il existe une approche différente pour mesurer la distance qui sépare deux vecteurs. Il suffit simplement de calculer la distance euclidienne entre les deux vecteurs:

$$dist = \|A - B\|^2.$$

Pour permettre l'invariance aux transformations linéaires, la moyenne des vecteurs, telle que définie dans l'équation 2.10, est soustraite des pixels de la région de corrélation et les vecteurs sont normalisés afin qu'ils soient unitaires. La forme normalisée de la mesure de distance euclidienne est définie comme:

$$\mathrm{ZNED}_{mn}(p_1, p_2) = \frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m}[(I_1(u_1+i, v_1+j) - \mu_1) - (I_2(u_2+i, v_2+j) - \mu_2)]^2}{V_1(u_1, v_1)V_2(u_2, v_2)} \tag{2.12}$$

où

$$V_k(u_k, v_k) = \sqrt{\sum_{i=-n}^{n}\sum_{j=-m}^{m}(I_k(u_k+i, v_k+j) - \mu_k)^2}.$$

Cette façon d'évaluer la distance entre deux vecteurs peut sembler simple et logique. D'ailleur, certaines méthodes de mise en correspondance l'utilise avec succès. La norme euclidienne a été utilisée dans les tests de la section 5.1.2. Les résultats

de la norme euclidienne et la corrélation normalisée sont montrées dans la figure 5.8 pour être comparés avec les résultats de la mesure ZNCC, montrées dans la figure 5.7. On peut constater que les résultats sont nettement supérieurs avec la mesure ZNCC.

2.2.2 Appariement dense

Afin de reconstruire un modèle de la scène basé sur des images, une façon apparemment simple serait d'obtenir la correspondance entre chaque pixel de chaque image. Un tel processus est appelé appariement dense et consiste en l'établissement d'un lien entre chaque pixel d'une image à un pixel correspondant dans l'autre image.

Le résultat de l'appariement dense est appelé carte de disparité. Il s'agit d'un champ de vecteur bidimensionnel qui fait correspondre les coordonnées de chaque pixel de l'image I_1 aux pixels de l'image I_2. Une carte de disparité peut être vue comme deux images, une image de déplacements horizontaux D_u et une image de déplacements verticaux D_v. D'où la relation entre les coordonnées d'un point $p_1 = (u_1, v_1)$ dans I_1 et d'un point $p_2 = (u_2, v_2)$ dans I_2 qui s'exprime comme suit:

$$u_2 = u_1 + D_u(u_1, v_1)$$

et

$$v_2 = v_1 + D_v(u_1, v_1).$$

Étant donné qu'une image contient un nombre élevé de pixels, faire de l'appariement dense est une tâche difficile, voir impossible, sans l'utilisation de supposition ou de connaissance préalable de la scène observée. Il est souvent nécessaire d'utiliser plus d'une contrainte afin de réduire la taille de l'espace de recherche. Dans les sections suivantes, les techniques d'appariement stéréo et de propagation de correspondance

seront décrites.

2.2.2.1 Mise en correspondance stéréo

Une bonne façon d'obtenir un appariement dense est d'utiliser des suppositions sur le mouvement des caméras entre les deux images. Lorsque l'on parle de mise en correspondance stéréo, les deux images sont prises par des caméras dont la position l'une par rapport à l'autre est connue. Lorsque les caméras sont calibrées, la matrice fondamentale peut être calculée en utilisant la transformation rigide entre les deux caméras tel que montré dans l'équation 2.3. Une fois que la matrice fondamentale est connue et que les images sont rectifiées pour avoir le même facteur d'échelle et orientation, la recherche, dans l'image I_2, pour la correspondance d'un point p_1 dans l'image I_1 se fait le long de la ligne épipolaire $l_2 = Fp_1$. Donc, la recherche est contrainte à un espace à une dimension le long de la ligne épipolaire générée par le point candidat dans l'image source (I_1) tel qu'illustré dans la figure 2.4.

Une façon pratique d'accomplir l'appariement stéréo est d'appliquer préalablement une transformation sur les images afin de rendre leurs lignes épipolaires parallèles. Dans (Seitz et Dyer, 1996*b*), on montre comment utiliser la matrice fondamentale F entre deux vues pour choisir deux transformations projectives (aussi appelées homographies) H_1 et H_2 telles que $H_2^T F H_1 = \hat{F}$, où $\hat{F}$ est la matrice fondamentale désirée entre les deux images rectifiées. Par exemple, la matrice fondamentale entre deux images dont les lignes épipolaires sont horizontales et parallèles est de la forme:

$$\hat{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

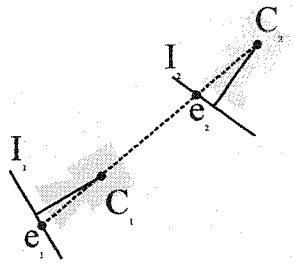


Figure 2.7 : Vues singulières.

La méthode décrite dans (Seitz et Dyer, 1996*b*) fonctionne pour des vues qui ne sont pas singulières. Une vue non singulière est telle que les épipôles (l'intersection des lignes épipolaires d'une image) ne sont pas à l'intérieur des images elles-mêmes. L'épipôle d'une image prise par une caméra représente la projection du centre optique de l'autre caméra. Une vue singulière est donc une vue dans laquelle on peut voir l'autre caméra. La figure 2.7 illustre un exemple de vues singulières. Les épipoles, qui correspondent à la projection du centre optique de l'autre caméra, sont à l'intérieur de l'image. Dans ce cas, les lignes épipolaires se croisent dans l'image et le processus de rectification va déformer l'image à un point tel qu'il sera impossible de l'utiliser par la suite pour faire la mise en correspondance.

Une fois que les images sont rectifiées, les lignes épipolaires correspondantes entre les deux images peuvent être mises en correspondance. Une façon d'accomplir ceci serait de trouver, pour chaque point dans la première image, le point correspondant dans la seconde image en appliquant une mesure de corrélation sur une fenêtre d'une certaine taille et trouver le point où la corrélation est maximum. Toutefois, quelques problèmes peuvent survenir.

Le premier est que, s'il n'y a pas assez de variations dans le signal par rapport au bruit, la corrélation ne va pas manifester un maximum clair et précis. Pour résoudre ce problème, la taille et la forme de la fenêtre de corrélation doivent être soigneusement

choisies selon la sorte de signal qui est mis en correspondance. La meilleure solution est d'avoir une fenêtre de taille variable qui grandit si le signal n'a pas ou très peu de variations et rapetisse si le signal a beaucoup de variations. Un tel algorithme est décrit dans (Kanade et Okutomi, 1992).

Un autre problème qui survient est qu'un patron peut être présent plus d'une fois dans le signal de l'image. Par conséquent, pour un patron de pixels dans la première image, plus d'un patron de pixels correspondants peut être trouvé dans la seconde image. Pour des scènes qui n'ont pas d'occlusions du tout, il est possible d'utiliser l'ordre, de gauche à droite, pour établir la correspondance entre les candidats possibles. Par contre, pour les scènes qui ont potentiellement des occlusions, d'autres contraintes doivent être utilisées, telle que la contrainte de régularité de la disparité. Le critère de régularité implique que pour deux points de correspondance voisins, la disparité doit varier de façon lisse et continue. Ce critère est souvent violé dans une scène réelle avec des arêtes droites et des occlusions. Il faut donc utiliser soigneusement ce critère en conjonction avec un algorithme de relaxation qui va préserver les discontinuités dans la carte de disparité.

L'appariement stéréo dense n'est vraiment utile que lorsque les caméras sont fortement calibrées, c'est-à-dire, lorsque les paramètres externes (rotation, translation) et internes (distance focale, distorsion des lentilles) sont connus avec précision. De plus, les algorithmes de correspondance stéréo supposent presque toujours une limite sur le déplacement entre les caméras, ceci afin de limiter la distorsion créée par les projections perspectives qui contribuent à fausser la corrélation sur des fenêtres rectangulaires.

2.2.2.2 Croissance de région

Au lieu de tenter directement d'obtenir une mise en correspondance dense, il peut s'avérer avantageux de tenter premièrement d'obtenir un appariement épars et ensuite de le raffiner afin d'obtenir un appariement dense. Il existe une méthode que l'on appelle croissance de région (*"region growing"*, en anglais) qui consiste à faire croître des régions de mise en correspondance dense autour de points d'appariement épars préalablement obtenus avec une autre méthode. Une telle méthode est décrite dans (Lhuillier, 1999). Cette méthode a été implantée au cours de ce travail afin d'obtenir une méthode éventuellement utile pour l'algorithme d'affinement de la triangulation. La figure 2.8 montre quelques résultats de la méthode de croissance de région sur une scène qui contient des roches. Il est intéressant de noter comment la frontière de l'occlusion (la roche en avant plan) est préservée. Dans cette expérience, les points de correspondance initiaux ont été produits à la main et ont été distribués uniformément dans les zones rocheuses de l'image.

La limite évidente de cette méthode de mise en correspondance est qu'elle a besoin d'une autre méthode pour obtenir l'appariement épars préalable. Par contre, dans la section 2.2.3, il est expliqué que l'obtention des points épars n'est pas aussi difficile que l'obtention de correspondance dense. Une limite plus importante de cette méthode est le manque d'appariement dans les régions de l'image où le signal varie peu. Ceci produit une carte de disparité avec beaucoup de régions non appariées. En ce qui concerne la synthèse de vues, ce genre de carte de disparité est insuffisant. Une méthode pour remplir les zones non appariées est nécessaire afin de produire des images sans trou.

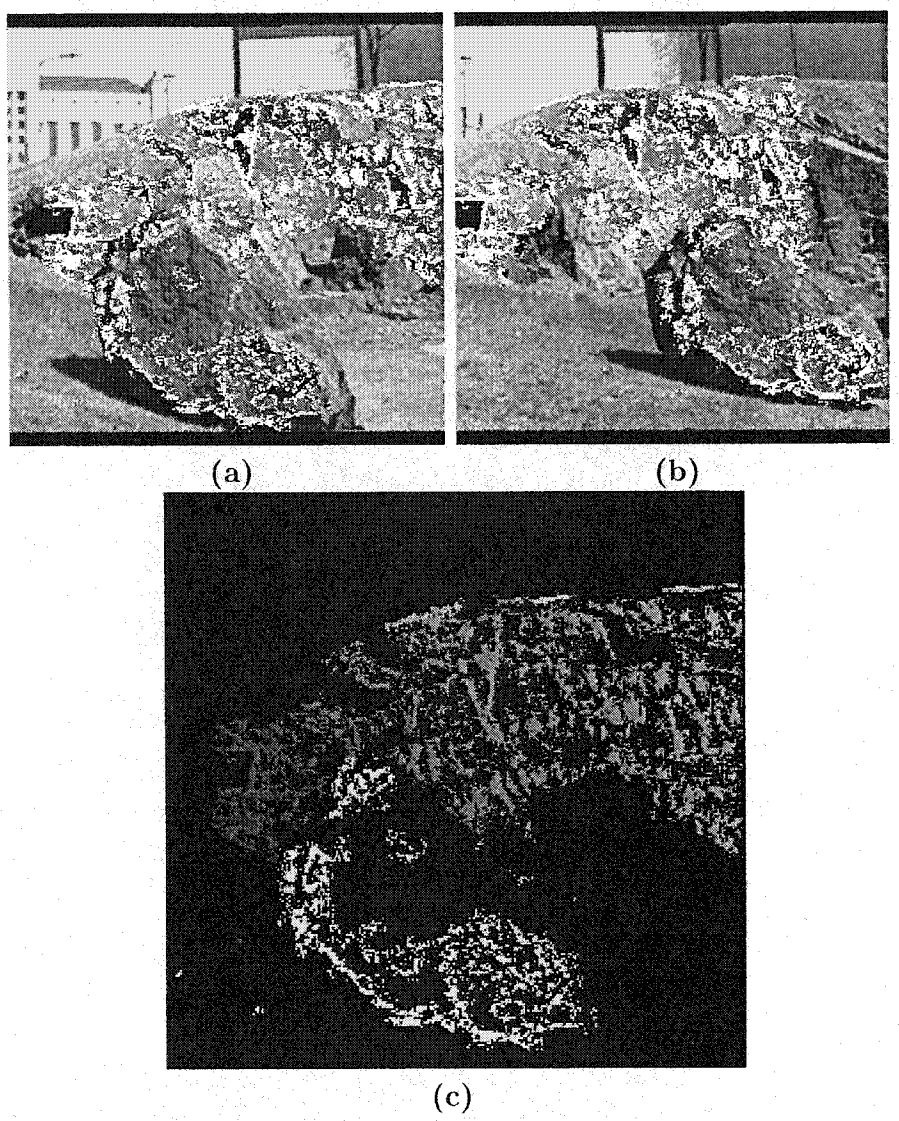


Figure 2.8 : **(a)** et **(b)** Images originales avec les points d'appariement (en blanc). **(c)** Carte de disparité. Les niveaux de gris indique la longueur des vecteurs de disparité.

2.2.3 Mise en correspondance éparse

La mise en correspondance éparse consiste à établir l'appariement d'un nombre limité de pixels isolés distribués de façon éparse dans l'image. Contrairement à l'appariement dense, l'appariement épars cherche à établir des correspondances uniquement aux endroits où les images contiennent des points d'intérêt. Il s'ensuit que la première étape de cette méthode est d'extraire les points d'intérêt contenus dans les images sources.

2.2.3.1 Détection des points d'intérêt

Un point d'intérêt dans une image doit avoir les caractéristiques suivantes: précision de localisation et stabilité. La précision de la localisation signifie que le point en question doit être détecté le plus près possible de sa vraie position dans l'image. La stabilité d'un point implique que le point détecté doit rester au même endroit dans différentes images de la même scène. Un type de point d'intérêt qui a le potentiel de respecter ces caractéristiques est ce que l'on appelle le coin. Un coin est habituellement défini comme un point de courbure planaire maximal dans la ligne de plus grande variation de la pente dans l'image en niveau de gris.

Beaucoup de recherche a été faite dans le domaine de la détection de coins. Une revue plus complète des méthodes de détection de coins peut être trouvée dans (Zheng, Wang et Teoh, 1999). Le reste de cette section va se concentrer sur l'explication d'un de ces détecteurs qui a été utilisé dans la méthode d'affinement de la triangulation décrite à la section 4.4.2. Il s'agit d'une mesure de la variation de la direction du gradient de l'image, définie comme:

$$\Delta(x,y) = \|\nabla\theta(x,y)\|^2 \cdot \|\nabla I(x,y)\|^2 - K\|\nabla I(x,y)\|^2 \tag{2.13}$$

où $I(x,y)$ est l'intensité de l'image en niveau de gris au point (x,y), $\nabla I(x,y)$ est le vecteur gradient de l'image, $\theta(x,y)$ dénote la direction du gradient au point (x,y) et, finalement, K est la constante de suppression de la réponse aux faux coins. Le vecteur gradient $\nabla I(x,y)$ est obtenu en faisant la dérivation partielle de premier ordre de l'image selon les deux axes:

$$\nabla I(x,y) = (\frac{\delta I}{\delta x}, \frac{\delta I}{\delta y}) = (I_x, I_y).$$

La direction du gradient $\theta(x,y)$ au point (x,y) est donnée par $\tan(\theta) = I_y/I_x$. D'où, $\theta(x,y) = \arctan(I_y/I_x)$. Le vecteur de variation de la direction du gradient $\nabla\theta(x,y) = (\delta\theta/\delta x, \delta\theta/\delta y)$, peut également être écrit comme $\nabla\theta(x,y) = (\theta_x, \theta_y)$.Une fois que toutes ces valeurs sont calculées, l'expression de $\Delta(x,y)$ peut être évaluée pour tout point (x,y). Les coins sont détectés en cherchant les maximums locaux lorsque $\Delta(x,y) > 0$.

Le premier terme de l'équation 2.13 mesure la quantité de variation dans la direction du gradient de l'image modulé par la norme du gradient. Pour qu'un coin soit détecté, il faut que la norme du gradient et sa variation angulaire soient suffisamment élevées. Le second terme de l'équation est utilisé pour prévenir la détection de faux coins, c'est-à-dire, lorsque la norme du gradient est grande, mais que la variation angulaire n'est pas suffisamment élevée. La constante K détermine quand la variation dans la direction du gradient est suffisamment élevée pour être considérée comme un coin. La valeur de K doit varier selon les masques de dérivation et de convolution gaussienne. Ceci introduit un effet arbitraire dans le choix du paramètre K. Il est plutôt proposé d'utiliser une fonction $K(x,y)$ qui est définie comme la convolution de l'opérateur gaussien $G(\sigma,x,y)$ avec la mesure de coin originale, dénotée comme

$\Delta_0(x,y) = \|\nabla\theta(x,y)\|^2$. Donc, $K(x,y)$ est défini comme:

$$K(x,y) = G(\sigma,x,y) \otimes \Delta_0(x,y).$$

Finalement, la mesure de coin est définie comme suit:

$$\Delta(x,y) = \|\nabla\theta(x,y)\|^2 \cdot \|\nabla I(x,y)\|^2 - K(x,y)\|\nabla I(x,y)\|^2 \tag{2.14}$$

Cette méthode de détection de coin est utilisée dans ce travail parce qu'elle a une meilleure précision de localisation que les autres détecteurs comme les opérateurs de Kitchen ou Plessey. Par contre, ce détecteur de coin à une performance de détection plus basse que Kitchen ou Plessey. Cependant, dans le cas de la mise en correspondance, il est plus important de trouver les coins à la bonne position que d'en trouver un plus grand nombre incorrectement positionnés.

2.2.3.2 Relaxation

Les points d'intérêts des deux images sources sont tout d'abord extraits. On obtient ainsi un ensemble de points ξ_1 provenant de l'image I_1 ainsi qu'un ensemble de points ξ_2 provenant de l'image I_2. Étant donné un point p_1 de l'ensemble ξ_1, un sous-ensemble de ξ_2 peut être choisi, soit en supposant une fenêtre de disparité maximum, par exemple un quart de la largeur de l'image, soit en utilisant la contrainte épipolaire si elle est connue. Le point p_1 est donc corrélé avec tous les points du sous-ensemble choisi dans l'image I_2. Les appariements candidats sont enregistrés lorsque le résultat de la corrélation est supérieure ou égale à une limite τ prédéfinie. En répétant ce processus pour tous les points de l'ensemble ξ_1, on obtient, pour chaque point, des candidats possibles pour l'appariement. Une technique de relaxation est ensuite util-

isée pour éliminer les appariements qui ne se correspondent pas, selon le critère de relaxation, jusqu'à ce qu'il ne reste plus qu'un seul candidat par point. Dans (Zhang et al., 1994), on décrit une méthode de relaxation qui attribue des énergies (ou des poids) de correspondance à chaque appariement potentiel. Ensuite, une itération de mise à jour permet d'éliminer les appariements potentiels les plus faibles. L'énergie de chaque appariement est ensuite recalculée pour l'itération suivante. Le processus se termine lorsque l'énergie des appariements ne varie plus, c'est-à-dire que le nombre de liens ne change plus.

2.2.4 Gérer les régions non appariées

Dans un cas idéal de mise en correspondance, chaque point (ou pixel) de l'image I_1 a un pixel correspondant dans l'image I_2. Alors, la synthèse de vues est faite simplement en reprojetant tous les pixels appariés des images dans la nouvelle vue. Ceci rend effectivement le processus de synthèse de vues indépendant de la complexité de la scène puisque le modèle de la scène est constitué de la carte de disparité. Aucun effort n'est fait pour établir un modèle compact et éviter la redondance d'information. C'est la méthode de force brute. En pratique, il est impossible d'obtenir une carte de disparité parfaite, c'est-à-dire sans zone non appariée. Il y a toujours des occlusions et des zones uniformes dans les images où il n'y a pas assez d'information dans la texture pour établir une correspondance. Même s'il était possible d'obtenir une correspondance dense sans lacunes, des trous seraient quand même créés dans l'image de destination lorsque les pixels sont reprojetés. Ceci se produit parce que les régions visibles de la scène vue par la nouvelle position de la caméra ne couvrent pas le même espace dans la nouvelle image. De plus, des régions de la scène qui étaient préalablement cachées peuvent devenir visibles et n'être pas couvertes par des pixels

reprojetés. Pour de petits trous, il existe quelques méthodes pour remplir les vides. Il est possible d'utiliser une génération de texture par échantillonnage randomisé dans la zone entourant le vide. On peut également remplir les trous avec la couleur des pixels voisins comme suggéré dans (Seitz et Dyer, 1996*b*). Toutefois, ces méthodes sont satisfaisantes seulement lorsque les trous sont très petits par rapport à la taille de l'image. Une méthode plus complète pour gérer les trous et les vides est donc nécessaire.

Une solution qui semble faire son chemin dans les nouveaux algorithmes de synthèse de vues est la triangulation des points de correspondance. Cette méthode implique de changer le modèle de la scène. Au lieu d'utiliser une carte de disparité, la scène est modélisée par un ensemble de points interconnectés pour former des triangles. Par exemple, dans (Lhuillier, 1999), on part d'une carte de disparité dense avec des trous pour la transformer en ce que l'auteur appelle une triangulation jointe des vues dont le résultat n'est autre qu'un ensemble de triangles qui couvre toute la surface de l'image et qui connecte les zones non appariées. Ensuite, pour produire de nouvelles images, les sommets des triangles sont reprojetés et remplis par application de textures affines ou en perspective, si la profondeur est connue.

2.3 Reprojection

Un concept de base en synthèse de vues est la reprojection des points qui proviennent des images sources sur la vue désirée. Il existe différentes méthodes pour accomplir cette opération. Une de ces méthodes consiste à utiliser les images sources pour faire une reconstruction 3D de la scène et ensuite, projeter cette scène dans les nouvelles images en utilisant les équations de projection perspective. Toutefois, d'autres méth-

odes accomplissent directement une reprojection, c'est-à-dire, en utilisant directement les coordonnées 2D des points dans les images sources pour générer des coordonnées 2D dans les images destinations.

Pour être plus précis, une reprojection est définie comme l'opération qui détermine la position d'un point dans une image destination à partir de la position des points correspondants dans deux ou plusieurs images sources ainsi que les paramètres qui définissent la nouvelle vue. On dénote les images sources I_1 et I_2 ainsi que la nouvelle image I_n qui correspond à un nouveau point de vue défini par l'intermédiaire des paramètres de la reprojection. Pour des points d'appariement donnés p_1 et p_2 dans I_1 et I_2 respectivement, la reprojection dans I_n sera notée p_n et est donnée par une fonction de reprojection $p_n = f(p_1, p_2, a, b, c, ...)$ où $a, b, c, ...$ sont les paramètres qui définissent le nouveau point de vue. Ces paramètres sont différents selon la méthode de reprojection utilisée.

Le reste de cette section se concentre sur la reprojection des points des images sources à l'image destination. La méthode de reprojection ne dépend pas de la méthode d'appariement par laquelle les points sont obtenus. Dans le cas d'appariement épars, une méthode de reprojection est utilisée pour les points, mais une façon de remplir le reste de l'image est nécessaire. Dans le chapitre 4, le processus de triangulation qui résous ce problème sera décrit.

Dans ce travail, les méthodes qui ont été utilisées pour la synthèse de vues ont été choisies pour leur simplicité et pour couvrir les types les plus communs de génération de vues. La première méthode, appelée interpolation de vues, est présentée dans la section 2.3.1 et la seconde méthode, appelée extrapolation de vues, est présentée dans la section 2.3.2.

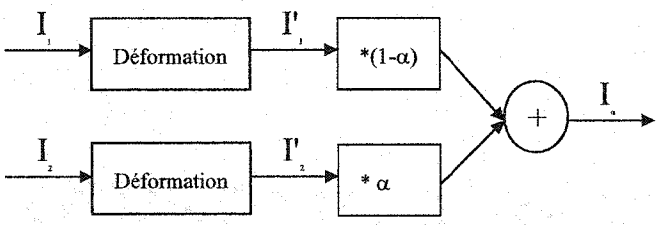


Figure 2.9 : Schéma du processus d'interpolation de vues.

2.3.1 Interpolation de vues

L'interpolation de vues est utilisée pour générer des images intermédiaires d'une scène. Cette technique est aussi parfois nommée *"view morphing"*, en anglais. Il s'agit de l'interpolation des points d'appariement entre les deux images selon la formule suivante:

$$p_\alpha = (1 - \alpha)p_1 + \alpha p_2 \tag{2.15}$$

où α est le facteur d'interpolation linéaire. La brillance du point résultant est aussi le résultat de l'interpolation des brillances des points sources afin de rendre la transition plus graduelle entre les deux images. En pratique, on implante cette méthode en générant, à partir des images sources I_1 et I_2, des images interpolées I'_1 et I'_2 selon l'équation 2.15. Ensuite, l'image finale est obtenue en combinant les deux images interpolées selon le facteur d'interpolation, c'est-à-dire que la brillance de chaque pixel de l'image finale est le résultat de l'interpolation linéaire de la brillance des pixels des images I'_1 et I'_2 à la position correspondante. La dissolution des images I'_1 et I'_2 utilise également le facteur d'interpolation α. La figure 2.9 illustre le processus d'interpolation de vues.

Pour certaines applications, cette méthode donne des résultats satisfaisants parce qu'il s'agit d'un problème bien posé en supposant que la scène est monotone (pas d'occlusions). Il a été montré que pour une configuration de caméras où les plans

images sont coplanaires, l'interpolation linéaire des points d'une image à l'autre est physiquement valide. Dans le cas où les plans images ne sont pas coplanaires, une rectification des points d'appariement est accomplie avant de faire l'interpolation. Ensuite, l'image résultante doit être "dérectifiée" pour obtenir l'image finale. Cette technique est décrite dans (Seitz et Dyer, 1996*b*).

2.3.2 Extrapolation de vues

On utilise le terme *extrapolation de vues* pour montrer la différence de concept avec l'*interpolation de vues* (section 2.3.1). L'interpolation est plus facile et plus fiable que l'extrapolation, mais elle est limitée par la position originale des caméras. D'un autre côté, l'extrapolation de l'information contenue dans les images permet une plus grande liberté de mouvement pour la caméra virtuelle. Il y a principalement deux approches à l'extrapolation de vue. La première approche consiste à faire une reconstruction 3D explicite et ensuite utiliser ces informations pour générer de nouvelles images. La seconde approche consiste à utiliser l'information 3D implicitement avec des invariants projectifs ou la contrainte épipolaire.

Dans ce travail, il a été plus utile de reconstruire la scène explicitement puisque la triangulation des points s'y prête bien. Une fois que les positions 3D des points d'appariement sont calculées, il ne reste qu'à projeter les triangles dans la nouvelle image et y appliquer la texture des images sources pour chaque triangle.

La partie difficile de cette approche est, évidemment, de calculer le modèle 3D de la scène. Le problème consiste à trouver la position dans l'espace d'un point P sachant la position de ses projections p_1 et p_2 dans les images I_1 et I_2 respectivement. Les sections qui suivent décrivent deux solutions à ce problème. Les deux solutions utilisent la matrice fondamentale (décrite dans la section 2.1) entre les deux images

sources.

2.3.2.1 Disparité stéréo

Cette méthode utilise la disparité stéréo pour estimer la profondeur des points. Afin d'évaluer la disparité stéréo des points, il faut tout d'abord rectifier les points dans les deux images sources. Le processus de rectification utilisé est décrit dans (Seitz et Dyer, 1996*b*) et plus en détail dans (Seitz, 1997). Pour résumer, la rectification applique une transformation projective (sous la forme d'une matrice 3×3 qui contient une homographie) aux points dans chaque images afin de les projeter sur un plan image commun. Les matrices d'homographies H_1 et H_2 qui vont amener les images sources sur un plan commun sont dérivées de la matrice fondamentale.

La transformation projective des points dans l'image I_j est décrite par les deux équations suivantes. Premièrement, la transformation du point par la matrice 3×3 H_j :

$$m_j = \begin{bmatrix} a_j \\ b_j \\ c_j \end{bmatrix} = H_j p_j \tag{2.16}$$

ensuite, le point est projeté sur le nouveau plan image:

$$p_j^r = \frac{m_j}{c_j} \tag{2.17}$$

où le point rectifié a la forme $p_j^r = [\ x_j^r \quad y_j^r \quad 1\ ]^T$.

Le but de cette rectification est de rendre les lignes épipolaires des deux images parallèles. Ainsi, chaque point correspondant se trouve sur la même ligne horizontale ou verticale. La disparité stéréo d peut ensuite être calculée à partir des points

rectifiés, p_1^r et p_2^r, comme suit:

$$d = \|x_1^r - x_2^r\| \tag{2.18}$$

Finalement, le point P dans l'espace est estimé en utilisant la disparité d:

$$P_j = \begin{bmatrix} X_j \\ Y_j \\ Z_j \end{bmatrix} = \frac{1}{d + f_d} \begin{bmatrix} x_j \\ y_j \\ f \end{bmatrix} \tag{2.19}$$

où x_j et y_j correspondent aux composantes du point $p_j = [\ x_j \quad y_j \quad 1\ ]^T$ avec $j = 1, 2$. f est la distance focale de la caméra "rectifiée" et f_d est une petit fraction de la distance focale utilisée pour prévenir les divisions par zéro. Cette valeur définit la profondeur maximale d'un point qui a une disparité égale à zéro.

Cette méthode ne génère pas un modèle de la scène qui est physiquement correct. L'estimation de la profondeur est relative à un plan commun calculé lors du processus de rectification. De plus, l'utilisation de la position (x, y) des points dans l'image est incorrect, mais dans certain cas, peut être satisfaisante pour des mouvements de caméra limités. Cette méthode permet donc d'estimer la profondeur des points dans les images.

2.3.2.2 Estimation du mouvement des caméras

Cette méthode consiste à retrouver le mouvement relatif entre les caméras. Une description plus détaillée de la technique peut être trouvée dans (Hartley, 1992). Pour simplifier le problème, on suppose que la distance focale et la position de la projection du centre optique de la caméra sont connues. En d'autres termes, on suppose que les paramètres internes de la caméra sont connus. On regroupe ces paramètres dans une

matrice afin que l'équation de projection puisse s'écrire de la façon suivante:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.20}$$

où f_x et f_y sont les facteurs de distorsion horizontale et verticale combinés avec la distance focale. La projection du centre optique est (x_0, y_0). On peut écrire l'équation dans une forme plus compacte comme suit:

$$p = \frac{1}{Z} AP \tag{2.21}$$

où A est la matrice des paramètres internes de la caméra.

Étant donné que A est connue, on peut l'enlever de l'équation en multipliant au préalable les points par A^{-1}. Les points ainsi normalisés sont calculés comme suit:

$$\tilde{p} = A^{-1} p \tag{2.22}$$

Cet ensemble de points normalisés peut être utilisé pour calculer la matrice fondamentale comme décrit à la section 2.1. Au lieu de calculer la matrice fondamentale, on obtient plutôt la matrice essentielle E. Cette matrice peut être décomposée de la façon suivante:

$$E = RT_S \tag{2.23}$$

où R est la matrice de rotation entre les deux caméras. T_S est la matrice oblique (*"skew matrix"* en anglais) du vecteur de translation T entre les deux centres optiques

des caméras. La matrice oblique d'un vecteur $T = [\ t_x \quad t_y \quad t_z\ ]^T$ est définie comme:

$$T_S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{2.24}$$

Pour récupérer la rotation et la translation relatives entre les deux caméras, il faut faire la factorisation de la matrice essentielle E en ces deux facteurs, la matrice orthogonale de rotation R et la matrice à symétrie oblique T_S. Pour accomplir ceci, il faut accomplir une décomposition en valeurs singulières:

$$E = UDV^T \tag{2.25}$$

où D est la matrice diagonale contenant les valeurs singulières. Le mouvement relatif est ensuite calculé comme une des solutions suivantes:

$$\begin{aligned} R = USV^T \quad &, \quad T = U[\ 0 \quad 0 \quad 1\ ]^T \\ R = USV^T \quad &, \quad T = -U[\ 0 \quad 0 \quad 1\ ]^T \\ R = US^TV^T \quad &, \quad T = U[\ 0 \quad 0 \quad 1\ ]^T \\ R = US^TV^T \quad &, \quad T = -U[\ 0 \quad 0 \quad 1\ ]^T \end{aligned} \tag{2.26}$$

où

$$S = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

et le vecteur de translation est normalisé, c'est-à-dire que $\|T\| = 1$.

Pour chacune des quatre solutions pour R et T, on peut calculer les points 3D. On définit le système de coordonnées afin qu'il coïncide avec le centre optique d'une des caméras. L'équation qui décrit la projection des points normalisés dans les images devient:

$$Z_2\tilde{p}_2 = Z_1 R\tilde{p}_1 + T \tag{2.27}$$

où les points $\tilde{p}_1$ et $\tilde{p}_2$ sont normalisés tel que décrit dans l'équation 2.22. Ensuite, si l'on écrit $R\tilde{p}_1 = [\ a \quad b \quad c\ ]^T$, on a le système d'équations suivant:

$$Z_2 \begin{bmatrix} \tilde{x}_2 \\ \tilde{y}_2 \\ 1 \end{bmatrix} - Z_1 \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{2.28}$$

où Z_1 et Z_2 sont les inconnues à résoudre. Ayant la profondeur de chaque point à partir de chaque plan image, on peut retrouver les points 3D puisque:

$$\tilde{P}_1 = Z_1\tilde{p}_1, \tilde{P}_2 = Z_2\tilde{p}_2$$

Les points retrouvés devraient faire face aux deux caméras, c'est-à-dire, que pour tous les points 3D calculés dans le système de coordonnées des deux caméras, il faut que $Z_1 \geq 0$ et $Z_2 \geq 0$. Seulement une des quatre solutions pour R et T va satisfaire cette condition pour tous les points et c'est cette solution qui est la bonne.

Cette méthode est meilleure que la méthode précédente qui utilise la disparité stéréo parce qu'elle permet d'obtenir la position des points, à un facteur d'échelle près, dans l'un ou l'autre des systèmes de coordonnées des caméras. Toutefois, il est difficile d'obtenir une matrice essentielle qui ne souffre pas de problème numérique.

Il faut que les paramètres internes soient connus avec précision. S'ils ne le sont pas, alors, les résultats obtenus sont très imprécis et souffre de problème de stabilité numérique.

2.4 Résumé sur la matière de ce chapitre

Ce chapitre a couvert les concepts et méthodes fondamentaux utilisés en synthèse de vues basée sur des images. Les principales méthodes de calibration et de mise en correspondance ont été présentées. Ces méthodes permettent de construire un modèle de la scène observée à partir des images sources. Une fois que le modèle de la scène est établi sous la forme requise pour l'application, une méthode de projection est utilisée pour synthétiser de nouvelles images de la scène. La méthode de projection peut utiliser une reconstruction tridimensionnelle explicite de la scène, dans ce cas, la synthèse de nouvelles images se fait par projection du modèle 3D de la scène selon le nouveau point de vue désiré. Toutefois, la reconstruction explicite d'un modèle de la scène est une tâche qui requiert une calibration précise des images utilisées. Obtenir une telle calibration n'est praticable que dans un environnement contrôlé. Si ce n'est pas le cas, alors un équipement de précision très coûteux est nécessaire. Heureusement, Il est possible de faire de la synthèse d'images sans reconstruire explicitement un modèle 3D de la scène. Le chapitre suivant traitera des différentes approches de synthèse de vues qui n'utilisent pas de reconstruction 3D explicite de la scène.

Chapitre 3

Revue des approches connexes

Comme il a été mentionné précédemment, la reconstruction tridimensionnelle explicite de la scène peut s'avérer difficile lorsque les conditions d'observation de la scène ne sont pas contrôlées parfaitement. Le but de ce travail est donc de proposer une méthode de représentation de la scène permettant une synthèse de vues sans reconstruction tridimensionnelle explicite. Pour concevoir une telle méthode, il est d'abord nécessaire de revoir les techniques de base pour la synthèse de vues qui ne font pas de reconstruction explicite du modèle 3D de la scène. La section 3.1 présente une telle revue. Ensuite, dans la section 3.2, d'autres méthodes pour établir un modèle polygonal physiquement valide de la scène sont survolées. Une revue plus complète et une classification des techniques de synthèse de vues basée sur des images peut être trouvée dans (Agam et al., 1999*b*).

3.1 Synthèse de vues à partir d'images faiblement calibrées ou non calibrées

Cette section contient un survol des méthodes de synthèse de vues qui n'utilisent que très peu ou pas du tout d'informations de calibration. Les méthodes sont classifiées selon les contraintes qu'elles imposent sur la position des points de vue qu'elles peuvent générer. Les approches de mosaïque d'images permettant de traiter différentes configurations de scènes sont revues dans la section 3.1.1. Les méthodes basées sur l'interpolation de vues pour générer des images intermédiaires sont présentées dans la section 3.1.2. Finalement, les approches qui permettent la génération d'images selon des points de vues arbitraires sont revues dans la section 3.1.3.

3.1.1 Approches basées sur des mosaïques d'images

La création d'une mosaïque se fait en collant deux ou plusieurs images ensemble pour créer une image plus grande ou avec une résolution supérieure (dans les zones où les images se superposent). Lorsqu'une mosaïque est créée à partir d'un groupe d'images, il est ensuite possible de générer de nouvelles vues de la scène selon certains points de vues limités. L'utilisation d'une mosaïque permet également de réduire l'espace requis pour entreposer les données puisque les images qui se superposent ne sont pas dupliquées. D'ailleurs, avant de faire la fusion des images, les zones de correspondance entre ces dernières doivent être établies. Plusieurs techniques utilisant différents types d'assemblage existent. Le type de mosaïque formé dépend des caractéristiques des images utilisées. Dans cette section, il sera traité des types de mosaïques plane, cylindrique et en couche. Une revue exhaustive des techniques de mosaïques est au-delà de la portée de ce document. Par contre, une telle revue peut être trouvée dans

(Kang, 1997).

3.1.1.1 Mosaïques de vues planes et cylindriques

Dans certaines applications, lorsque la variation de profondeur dans la scène est négligeable comparée à la distance de l'observateur, il est possible de considérer la scène comme une surface approximativement plane. Dans un tel cas, la complexité du problème est réduite puisque la distorsion due à la perspective est minime d'une image à l'autre. Une méthode d'assemblage de mosaïques d'images aériennes ou orbitales, qui satisfont la contrainte de planarité précédente, est décrite dans (Moffitt et Mikhail, 1980). Une fois que les images sont appariées, la mosaïque est créée en gardant la brillance des pixels d'une des images. Toutefois, ceci produit des bordures visibles entre les images. Des techniques variées, décrites dans (Milgram, 1975; Milgram, 1977; Peleg, 1981; Burt et Adelson, 1983), permettent de faire le lissage ou même la soustraction de ces bordures visibles. Ces techniques permettent d'avoir une mosaïque uniforme.

Il arrive toutefois que la distorsion perspective ne soit pas minimale et qu'il faille utiliser une autre approche. Si le mouvement de la caméra consiste en une simple rotation sur un pivot fixe situé au centre optique de la caméra, il est possible de créer une mosaïque avec de telles images. Dans ce cas, les images sont reliées par une transformation projective linéaire correspondant à une rotation. Ainsi, des mosaïques cylindriques et sphériques (Irani, Anandan et Hsu, 1995; Szeliski et Kang, 1995) peuvent être créées pour faire des vues panoramiques. Lorsque la mosaïque générée est représentée par une image rectilinéaire, le champ de vision est limité à environ 180°. Pour un champ de vision plus large, une représentation différente est nécessaire. Dans ce cas, des coordonnées cylindriques sont utilisées (Chen, 1995).

La reprojection d'un ensemble d'images sur une surface cylindrique commune peut être évitée lorsque le but de l'opération est de générer une vue panoramique complète à partir d'une séquence d'images, tel que décrit dans (Shum et Szeliski, 1997; Szeliski et Shum, 1997). L'avantage de cette approche est que le mouvement de la caméra n'a pas besoin d'être contrôlé. Il n'y a pas de contraintes sur la façon dont les images doivent être prises pourvu qu'il n'y ait pas de mouvements de parallaxe marqués. Cette approche est basée sur une représentation des images de la mosaïque comme un ensemble de transformations. Ainsi, les problèmes de singularité qui existent lors de l'usage de coordonnées sphériques ou cylindriques, au bas et au haut du cylindre ou de la sphère, sont évités.

L'utilisation de mosaïques panoramiques pour la réalité virtuelle est démontrée dans (Yong, Xuehui et Enhua, 1997; Szeliski, 1996). Un ensemble d'images est utilisé afin de composer un environnement virtuel. Une visite dans l'espace virtuel est ensuite obtenue en sautant d'un point de vue panoramique à un autre et en synthétisant de nouvelles images à partir de la mosaïque panoramique courante.

3.1.1.2 Mosaïques en couches

Un aspect plus récent de la technique d'assemblage en mosaïque est l'utilisation d'information de profondeur lors de la construction de la mosaïque. Dans cette approche, les images sont décomposées en couches dont la profondeur moyenne de leurs pixels diffèrent du reste de l'image. Chaque pixel d'une couche possède une couleur, une intensité et un niveau de transparence. Un exemple d'un tel système est décrit dans (Kamei, Maruyama et Seo, 1997). Les auteurs y présentent une méthode qui génère une nouvelle vue en la séparant en couches et en y collant dans chaque niveau de profondeur les couches les plus semblables parmi les images enregistrées préalable-

ment. Cette approche est capable de synthétiser des approximations adéquates de scènes complexes sans toutefois nécessiter une grande quantité d'images sources. Une représentation de mosaïque en couche peut être appliquée à des séquences d'images, tel que décrit dans (Adelson, 1995). Dans cette approche, les images d'une séquence sont mises en couches de façon à ce que chaque couche contienne un ensemble de cartes 2D recalées. Ces cartes 2d sont constituées de cartes de brillance, de transparence, de vitesse, de profondeur, perturbation, de variation, etc. Une image peut être affichée en faisant la composition des couches en ordre de profondeur. Cette représentation permet une analyse des mouvements et une segmentation améliorée.

Le processus de définition des couches peut être fait de façon hiérarchique, tel que décrit dans (Kumar, Anandan, Irani, Bergen et Hanna, 1995). Cette hiérarchie de représentations est donnée par: les images elles-mêmes, une mosaïque d'images 2D, une mosaïque d'images avec parallaxe et finalement, une mosaïque avec des couches et des tuiles avec parallaxe. En se basant sur la disposition de la scène et la géométrie des mouvements de la caméra, une de ses représentations préserve l'exactitude de la scène. Chaque représentation nécessite différents algorithmes pour être formée. Plus cette représentation se trouve haut placée dans la hiérarchie, plus elle est complexe, mais lorsqu'elle est formée, elle peut être utilisée pour générer de nouvelles vues de la scène à partir de points de vue intermédiaires.

3.1.2 Interpolation de vues

Lorsque le résultat de la synthèse d'une nouvelle vue est contraint d'être situé entre deux vues déjà existantes, des simplifications peuvent être appliquées. Ce cas particulier est appelé, dans la littérature, *interpolation de vues*. Les méthodes variées d'interpolations sont classifiées selon trois catégories : les approches basées sur la

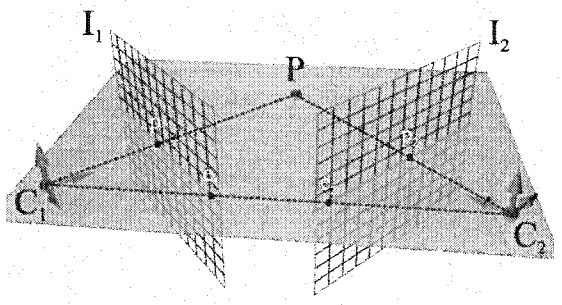


Figure 3.1 : Vue schématique de la géométrie épipolaire. Le plan épipolaire passe à travers le point P et les centres de projections c_1 et c_2. Les points p_1 et p_2 sont des projections du point p dans les deux plans images. Les points e_1 et e_2 sont les épipoles des images pour cette configuration.

géométrie épipolaire, les approches basées sur la combinaison linéaire de vues et les techniques de déformation de vue.

3.1.2.1 Images de plan épipolaire

Cette section couvre les approches de synthèse de vues qui sont basées sur la *géométrie épipolaire* de deux images (ou vues) (voir (Zhang, 1996) pour une introduction à la géométrie épipolaire ou encore la section 2.1). Tel qu'illustré dans la figure 3.1, un point P dans l'espace 3D est projeté sur les points p_1 et p_2 dans les plans images I_1 et I_2 respectivement. Les centres de projection c_1 et c_2, ensemble avec le point P forme un plan épipolaire. L'intersection de ce plan avec les plans images définit des lignes épipolaires dans les images. Dans une image, les lignes épipolaires se croisent toutes au même point. Ce point est appelé l'épipôle de l'image. Les épipôles e_1 et e_1 sont situés à l'intersection de la ligne qui joint les deux centres de projection, c_1 et c_2, avec les plans images. Donc, la ligne épipolaire dans l'image I_i passe à travers les points p_i et e_i.

Étant donné une configuration avec plusieurs points de vue, dans laquelle plusieurs caméras également séparées sont situées sur une ligne droite et dont les plans images

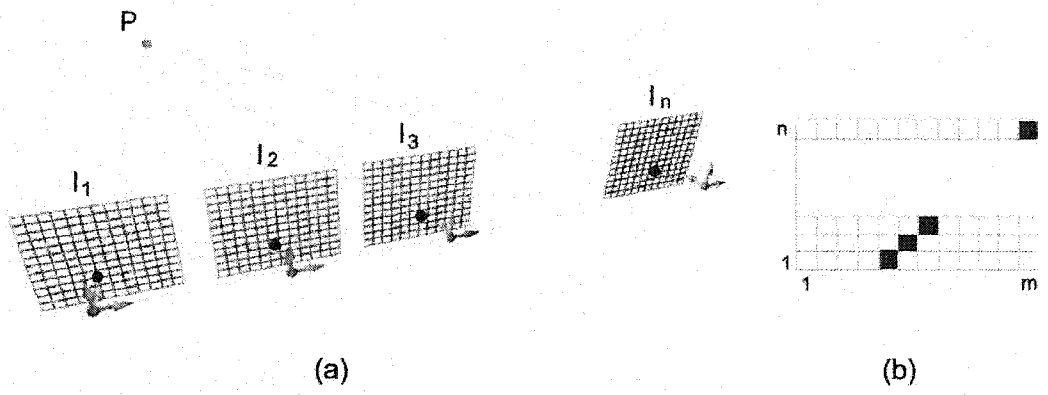


Figure 3.2 : Génération d'une image de plan épipolaire. Une IPE est créée en empilant les lignes épipolaires correspondantes prises d'une séquence d'images également séparées et où les plans images sont coplanaires et les axes des plans images sont alignés. **(a)** Une séquence de n vues représentées par les plans images $I_1 - I_n$. Le point P est projeté sur chaque plan image. **(b)** Une des IPE générée. L'IPE a n rangées de longueur m identique.

sont parallèles (Heung-Yeop, Jin-Ho, Je-Ho, Yong-moo, Sangkuk et Sang-Hui, 1997), il est possible d'accumuler les lignes épipolaires dans des images, appelées images de plan épipolaire. Chaque image de plan épipolaire (IPE) (Bolles, Baker et Marimont, 1987) est formé d'une pile ordonnée de lignes épipolaires correspondantes provenant des multiples points de vue. La génération d'une IPE est illustrée dans la figure 3.2. Un point 3D génère une droite dans l'IPE, où la pente de cette droite est reliée à la distance entre le point et la caméra. Les occlusions des points caractéristiques sont représentées, dans les IPE, par des lignes qui s'entrecroisent. La ligne qui a la plus petite pente correspond au point qui cache l'autre. En utilisant des IPE, la génération d'une vue intermédiaire consiste à rechercher les trajectoires linéaires des points caractéristiques et d'interpoler de nouveaux points le long de ces lignes. L'utilisation des IPE simplifie le problème de correspondance de vue et permet de déterminer et de gérer les occlusions dans la scène.

Une configuration avec plusieurs points de vue placés uniformément sur un chemin circulaire avec leur axe dirigé perpendiculairement à ce chemin est décrit dans (Hsu, Kodama et K., 1994). Dans ce cas, la trajectoire des points caractéristiques dans l'IPE forme des courbes périodiques modélisées par des splines cubiques. L'application des IPE sur une séquence d'image est décrite dans (Baker et Bolles, 1989). On y décrit comment empiler des IPE correspondantes l'une sur l'autre afin de former un cube que l'on appelle le volume de plan épipolaire (*"epipolar plane volume"*, EPV, en anglais).

3.1.2.2 Combinaison linéaire de vues

Une combinaison linéaire de vues existantes peut être utilisée pour générer de nouvelles vues intermédiaires de la scène. Cette approche nécessite d'abord l'établissement de correspondance entre les images existantes. Cette mise en correspondance peut être éparse (i.e. basée sur un petit nombre de points épars) ou dense. En partant de la correspondance établie, une interpolation linéaire définie par $p_\alpha = (1-\alpha)p_1 + \alpha p_2$ produit le point interpolé p_α dans la vue intermédiaire. Dans cette dernière expression, p_1 et p_2 représentent les points correspondants dans les vues existantes et $0 \leq \alpha \leq 1$ est le facteur d'interpolation. Tous les points interpolés sont ensuite utilisés pour produire une déformation (*"warp"* en anglais) des images existantes qui sont ensuite combinées par inter-dissolution. Le processus de déformation prend chacune des images sources et les déforme à l'aide des points de correspondance afin que ceux-ci se placent à leur position correspondante dans l'image interpolée. Ensuite, les images sources ainsi déformées sont combinées en une seule image par un processus d'inter-dissolution. Ceci correspond simplement à faire la combinaison linéaire selon le facteur d'interpolation α de chacun des pixels des images sources de la façon suivante: $I_\alpha(x,y) = (1-\alpha)I_1(x,y) + \alpha I_2(x,y)$ où $I_i(x,y)$ est l'intensité lumineuse de l'image

I_i au point (x, y). Cette approche est souvent appelée *déformation d'image*. Cependant, il faut prendre en note que l'image ainsi générée peut ne pas être physiquement correcte dans les cas où il y a des occlusions dans la scène ou si les plans images des vues ne sont pas coplanaires.

Un exemple de cette approche est présenté dans (Chevrier, 1997), où une interpolation linéaire de vues est utilisée comme une façon efficace de générer de nouvelles images d'un modèle 3D. Dans cette approche, un ensemble d'images clés du modèle 3D est préalablement calculé avec une méthode conventionnelle de lancé de rayons. Ensuite, cet ensemble d'image est utilisé pour générer de nouvelles vues par déformation des images clés. Des informations 3D supplémentaires permettent l'interpolation des caractéristiques spéculaires et diffuses des environnements sans avoir à estimer les mouvements de caméra. Une approche similaire est décrite dans (Werner, Hersch et Hlavac, 1995). On y présente une méthode de représentation d'objets basée sur une combinaison linéaire d'un groupe d'images 2D. Un modèle de caméra affine est utilisé. La qualité du processus d'interpolation est limitée par la distance entre les différentes vues. L'erreur introduite dans les images interpolées est proportionnelle à la distance entre les vues utilisées.

Il est possible d'utiliser la synthèse de vues par combinaison linéaire pour faire de la reconnaissance d'objet. Cette méthode consiste à utiliser un ensemble d'images de l'objet à reconnaître et à générer de nouvelles images de cet objet afin de les comparer avec une image réelle de l'objet. Un exemple de cette approche est présenté dans (Edelman et Bulthoff, 1992). On y décrit un système de reconnaissance 3D qui utilise cette technique pour modéliser la reconnaissance visuelle humaine. Dans ce système, les images synthétisées de vues caractéristiques des objets sont utilisées pour reconnaître ces derniers.

La technique d'inter-dissolution des images déformées, tel qu'expliqué précédemment, utilise habituellement la moyenne pondérée des pixels correspondants des images sources. Un des problèmes de cette méthode est que l'image produite peut être embrouillée si les images sources ne correspondent pas vraiment. Ceci peut arriver puisque le processus de déformation des images n'est pas exact et ne tient pas compte directement de la géométrie de la scène. Le travail décrit dans (Mansouri et Konrad, 1997) propose une solution pour régler ce problème. L'algorithme décrit utilise un filtrage non-linéaire qui applique une stratégie du gagnant-prend-tout plutôt que de faire une moyenne pondérée des pixels. On assume que l'image reconstruite est un grillage de tuiles de dimension fixe provenant de diverses positions dans les images de droites ou de gauches reliées par une carte de disparité. Le grillage de tuiles est modélisé par un champ de décision binaire alors que le modèle de disparité est basé sur la contrainte de continuité (*"smoothness"* en anglais).

Un autre problème fréquent avec la déformation d'images est la création de trous ou de plis dans l'image synthétisée lorsque l'information de mise en correspondance est incomplète ou insuffisante. Ce problème a été étudié dans (Fujimura et Makarov, 1998). Une méthode y est présentée pour déformer une image de façon continue sans faire de plis tout en respectant la trajectoire des points d'ancrage. Cette méthode est basée sur une triangulation chrono-dynamique, i.e. une triangulation qui se modifie selon le mouvement des points d'intérêt. Les points d'intérêt incluent également des segments de droite ou des polygones. Une méthode pour remplir les trous basée sur le voisinage local est décrite dans (Scharstein, 1996).

Finalement, le problème de la réduction de la résolution dans les images intermédiaires est étudié dans (Owen et Makedon, 1997). On y présente une méthode de déformation affine d'images qui à la particularité d'être séparable. Une matrice de

transformation affine est décomposée en trois opérations matricielles de base : une transformation oblique, une translation et une mise à l'échelle sur un seul axe.Cette décomposition est ensuite utilisée dans un algorithme à trois passes où chaque passe est constituée d'une seule de ces opérations.

3.1.2.3 Déformation physiquement correcte d'image

Une méthode améliorée d'interpolation de vues, basée sur la déformation d'images, combine l'interpolation des textures et de la géométrie de la scène. On appelle cette technique la déformation physiquement correcte d'images ou, simplement, *déformation de vues*. Les travaux décrits dans (Chen et Williams, 1993) utilisaient la déformation d'images afin d'éviter de recalculer de nouvelles images d'une scène complexe à partir d'un modèle 3D de cette scène. On utilisait un ensemble d'images pré-calculées du modèle ainsi que les cartes de profondeur associées. Une nouvelle image de la scène était créée en déformant les images sources et en utilisant leur carte de profondeur dans les paramètres de la déformation. Cette approche utilisait de l'information 3D qui devait d'abord être calculée. Il a été découvert plus tard, dans (Seitz, 1997), que le besoin d'information 3D n'était pas nécessaire pour ce type de synthèse de vues. En effet, l'information se retrouve implicitement dans la collection d'images et d'appariements entre celles-ci.

La déformation de vues a été étudiée en profondeur dans (Seitz et Dyer, 1995*b*; Seitz et Dyer, 1995*a*; Seitz et Dyer, 1996*b*; Seitz et Dyer, 1996*a*; Dyer, 1997; Seitz et Dyer, 1997). Cette approche est une extension de la déformation d'images qui gère correctement les caméras perspectives et les transformations de scène. L'avantage principal de cette technique est que les images n'ont pas besoin d'être fortement calibrées et, donc, que les images peuvent être traitées directement. Cette méthode

consiste à produire deux images corrigées en leur appliquant une déformation préalable de façon à ce que leurs lignes épipolaires deviennent mutuellement parallèles et alignées. Ensuite, la déformation est faite sur les images corrigées. Le résultat de cette opération est ensuite déformé à nouveau de façon à défaire la déformation préalable appliquée aux images sources. La pré-déformation, qui est en fait une reprojection perspective plane (homographie), est déterminée par la géométrie épipolaire. Cette dernière doit être calculée à partir d'information de correspondances entre les images, ou bien en utilisant des étalons de calibration. Pendant le processus de déformation, une mise en correspondance dense est établie entre les pixels de chaque ligne épipolaire afin de produire la ligne épipolaire interpolée. Tel que mentionné plus tôt, cette technique ne nécessite pas d'information 3D à priori et les images générées sont physiquement correctes sous l'hypothèse que la scène est monotone. Ceci implique qu'il n'y a pas d'occlusions dans la scène selon les points de vue utilisés. La post-rectification de l'image interpolée, qui a pour but de corriger la distorsion introduite par la pré-déformation, doit être spécifiée de façon interactive par l'utilisateur. Le processus de déformation de vues est illustré dans la figure 3.3.

Finalement, une technique de déformation de vues dynamique est présentée dans (Manning et Dyer, 1998). On suppose que la scène observée est constituée d'un petit nombre d'objets qui peuvent subir une translation entre les vues.

3.1.3 Génération d'images à partir d'un point de vue arbitraire

L'interpolation de vues ne permet de générer que des images intermédiaires entre deux images déjà existantes. Toutefois, il peut être nécessaire, dans certains cas, de

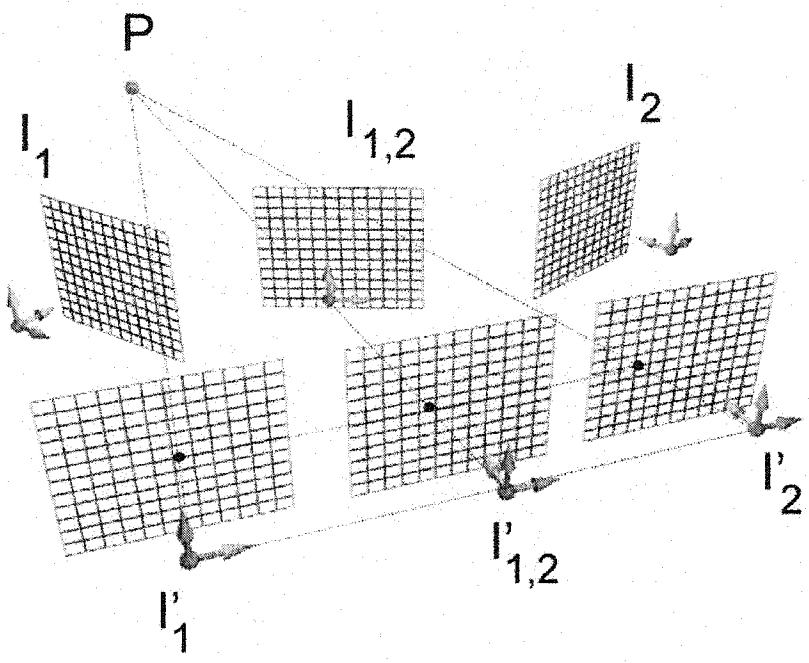


Figure 3.3 : Illustration du processus de déformation de vues. Les images pré-déformées $I_1^{'}$ et $I_2^{'}$ sont générées à partir des images I_1 et I_2 en les projetant sur un plan image commun qui va rendre leurs lignes épipolaires alignées. L'image $I_{1,2}^{'}$ est produite par la déformation des images $I_1^{'}$ et $I_2^{'}$. Finalement, la vue synthétisée $I_{1,2}$ est générée par la post-rectification $I_{1,2}^{'}$.

générer de nouvelles vues à partir de point de vue arbitraire. Ces points de vues sont quand même contraints à certaines positions d'intérêt puisque les images sources ne contiennent pas nécessairement assez d'information pour permettre la synthèse d'images sans contraintes. Il s'agit de contraintes similaires à celles que l'on retrouve dans les techniques de reconstruction de modèle 3D de la scène qui permettent, elles aussi, de synthétiser des images de point de vue arbitraire.

3.1.3.1 Transformations géométriques 2D

Lorsque les transformations géométriques entre des images sont contraintes à des cas simples, la synthèse de vues peut être faite directement à partir des plans image bidimensionnels. Un exemple d'un tel cas est présenté dans (Park, Yagi et Enami, 1994). On y observe que toute opération de caméra contenant un changement de mise à l'échelle et une rotation 3D pure peut être représentée par une transformation géométrique 2D. On peut ensuite utiliser cette transformation pour produire de nouvelles images. Afin d'y parvenir, on estime les paramètres de la caméra à partir d'une séquence d'images. Ce processus est basé sur l'établissement de correspondances qui tiennent compte de plusieurs contraintes : l'angle de vue, le changement d'échelle et la direction de la caméra. Une approche similaire est présentée dans (Ullman et Basri, 1991) où on utilise une telle méthode pour la reconnaissance d'objets.

3.1.3.2 Invariants projectifs

Une approche plus générale pour générer de nouvelles vues à partir d'une position arbitraire est décrite dans (Laveau et Faugeras, 1994). Les données sources utilisées sont des vues calibrées, soit faiblement, par résolution de la géométrie épipolaire ou fortement, en utilisant un étalon de calibration. Dans cette approche, la scène est

représentée par une collection d'images reliées entre elles par des matrices fondamentales exprimant la géométrie épipolaire de chaque paire de vues. Cette représentation permet d'éviter une reconstruction 3D explicite de la scène. Étant donné deux vues de la scène, l'utilisateur doit spécifier interactivement le centre optique et le plan rétinien de la caméra virtuelle désirée dans chaque image source. Cette technique demande une connaissance des matrices fondamentales entre les images ainsi qu'un appariement dense. Les auteurs décrivent ensuite comment générer la nouvelle image en utilisant la contrainte épipolaire pour trouver l'intersection des lignes épipolaires, à un facteur d'échelle près. Si le système est fortement calibré, le facteur d'échelle est déterminé. Cette approche de base est ensuite étendue pour faire usage de vues multiples.

Une approche différente, qui utilise de l'appariement épars, est décrite dans (Havaldar, Lee et Medioni, 1996; Havaldar, Lee et Medioni, 1997; Chen et Medioni, 1997). Les points d'intérêts sont détectés et appariés manuellement dans les deux vues. Ces points d'intérêts comprennent également des lignes qui représentent les arêtes des objets dans la scène. Une triangulation contrainte qui préserve ces arêtes est calculée sur chaque image. On assume donc que les triangles produits correspondent à des surfaces d'objet dans la scène et que deux triangles correspondants sont reliés par une homographie. Cette dernière peut être obtenue en utilisant les trois points correspondants des triangles et les épipôles des images comme quatrième point. Afin de déterminer la position des triangles dans la nouvelle image, on utilise la profondeur projective qui est basée sur le calcul d'un invariant appelé produit croisé (*"cross ratio invariant"*, en anglais). Les triangles dans la nouvelle image sont ensuite dessinés par plaquage de textures. Afin de résoudre les occlusions, les triangles sont préalablement triés et dessinés selon l'ordre de profondeur projective. Cette approche permet égale-

ment de générer des images à partir de plus de deux images. Une autre approche, décrite dans (Barrett, Payton et Marra, 1997), utilise des relations invariantes entre les images de références pour transférer les points et les lignes conjugués dans la vue synthétisée. Ces relations invariantes sont estimées de façon empirique.

3.1.3.3 Tenseur trilinéaire

La matrice fondamentale définie la contrainte linéaire qui relie deux vues. Il existe également une relation entre trois images. On appelle cette relation le *tenseur trilinéaire* (Shashua, 1995; Shashua, 1997). Dans ce formalisme, les contraintes linéaires reliant trois différentes vues peuvent être représentées par un tenseur à trois indices α^{ijk} qui relie chaque point des trois images. Tout comme la matrice fondamentale, il est possible de calculer le tenseur à partir de sept points de mise en correspondance entre les trois images. L'avantage obtenu par ce formalisme est de permettre la génération de nouvelles vues à partir de deux images densément appariées ainsi qu'une troisième image ayant au moins sept points d'appariement avec les deux premières. Des approches numériques pour calculer le tenseur trilinéaire sont présentées dans (Faugeras et Papadopoulo, 1997). Les contraintes linéaires qui relient les trois vues sont illustrées dans la figure 3.4, où $P = (X, Y, Z)^T$ est un point dans l'espace 3D qui est projeté sur les trois plans images différents, ce qui donne naissance aux points 2D p_1, p_2 et p_3, respectivement. Le tenseur trilinéaire exprime les contraintes sur les lignes qui passent par p_2 et p_3, étant donné un point p_1 dans la première image.

L'application du tenseur trilinéaire pour la synthèse de vues est décrite dans (Avidan et Shashua, 1997*a*; Avidan, Evgeniou, Shashua et Poggio, 1997). Dans cet article, deux images densément appariées et une troisième image avec sept points d'appariement sont utilisées pour générer un tenseur source $\alpha_{<123>}$. La translation

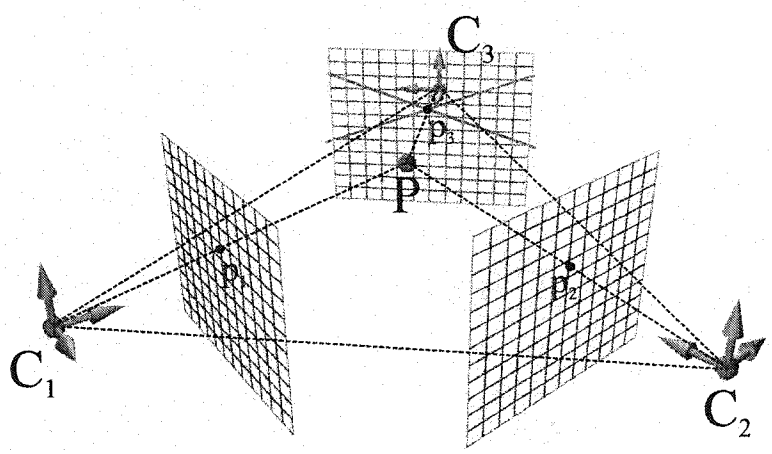


Figure 3.4 : Illustration de la relation entre les lignes et les points dans les trois plans images. Cette relation est exprimée dans le tenseur trilinéaire. Les points p_1–p_3 sont les projections du point 3D P dans les trois plans images où c_1–c_3 sont les centres de projection. Les lignes dans la troisième image représentent les lignes épipolaires.

(**T**) et la rotation (**R**) qui amènent un des points de vue existant à une nouvelle position sont spécifiées par l'utilisateur et sont utilisées avec le tenseur source $\alpha_{<123>}$ pour générer le tenseur $\alpha_{<12\psi>}$ qui correspond à la nouvelle configuration de vues. Ce tenseur est ensuite utilisé pour générer la nouvelle vue.

Tel que noté dans (Avidan et Shashua, 1997*b*), le formalisme tensoriel peut être appliqué dans le cas dégénéré de deux vues où les contraintes trilinéaires dégénèrent en contraintes bilinéaires. Dans ce cas, la théorie sous-jacente est essentiellement celle de la matrice fondamentale telle que décrite dans la section 3.1.2.1.

3.2 Aspects additionnels de la synthèse de vues

Cette section discute de certains aspects additionnels de la synthèse de vues qui n'ont pas été couverts jusqu'à maintenant. Étant donné un groupe d'images reliées par correspondance éparse, plusieurs méthodes de synthèse de vues utilisent une triangulation des points d'appariement. Le problème qui consiste à obtenir une triangulation

correcte est abordé dans la section 3.2.1.

3.2.1 Triangulation correcte

Le but d'une triangulation correcte est de produire des triangles qui correspondent parfaitement à des surfaces planes dans la scène. Cette approche est appelée *triangulation physiquement valide (TPV)*. Une TPV peut ensuite être utilisée pour générer de nouvelles vues de la scène à l'aide d'une des méthodes de synthèse de vues présentées précédemment. Il est à noter que la connexion des points pour former des triangles dans une vue implique la même connexion dans une seconde vue, ainsi, la procédure est aussi appelé *triangulation de vues jointes* (*"Joint View Triangulation"*, en anglais).

La TPV est importante autant pour les vues densément appariées que pour les vues à correspondances éparses. Puisque dans chaque cas, il y a invariablement des parties des images qui ne sont pas mises en correspondance et donc, qui doivent de toute façon se fier à des points épars. La triangulation des points épars présente une solution possible pour la gestion des régions non-appariées. De plus, la synthèse de vues utilisant une TPV offre une complexité de calcul diminuée et peut être accélérée par les cartes graphiques courantes. La production d'une TPV nécessite une interprétation de la scène. De plus, des erreurs dans la triangulation peuvent conduire à des artefacts significatifs dans les images synthétisées. La figure 3.5 illustre la triangulation des points d'ancrage dans l'image d'un cube. La figure 3.5-a présente la vue d'origine avec les sept points d'appariement (A–G). La figure 3.5-b montre une triangulation physiquement invalide (T_{FCA}, T_{FAD}, T_{FDE}, T_{FEG}, T_{DAB}, T_{DBE}). La figure 3.5-c montre une TPV (T_{DCA}, T_{DAB}, T_{DBE}, T_{DEG}, T_{DGF}, T_{DFC}) dans laquelle tous les triangles correspondent à des surfaces planes dans la scène.

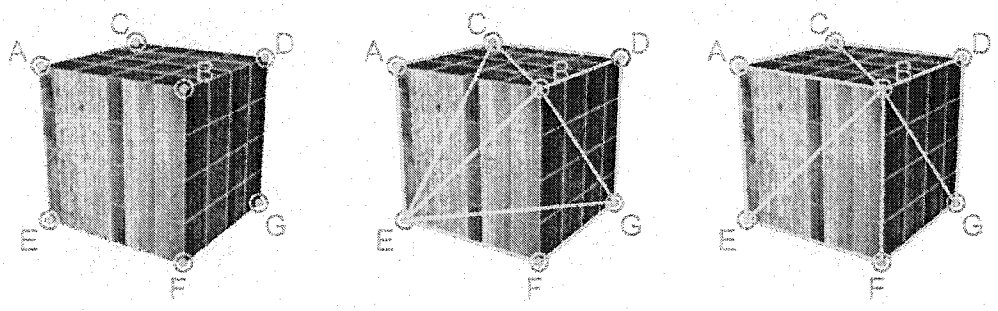


Figure 3.5 : **(a)** Une image d'un cube avec les points d'appariements. **(b)** Triangulation physiquement invalide dans laquelle certaines arêtes des triangles croisent les faces du cube. **(c)** Triangulation physiquement valide dans laquelle tous les triangles appartiennent à des surfaces du cube.

On peut obtenir une triangulation valide en imposant des contraintes sur le processus de triangulation. Une telle contrainte est d'empêcher que les arêtes des triangles croisent des arêtes dans l'image. Ceci fait l'hypothèse que les arêtes dans l'image correspondent bien à des arêtes dans la scène. Dans (Havaldar et al., 1996; Chen et Medioni, 1997) (voir aussi la section 3.1.3.2), on utilise une triangulation de Delaunay contrainte. Dans ce cas, la validité de la triangulation obtenue dépend de la validité et complétude de l'identification, dans l'image, des arêtes de la scène. Toutefois, puisque qu'un algorithme automatique ne peut pas faire la différence entre une arête dans la texture d'un objet et une arête qui correspond à la jonction entre deux objets différents, cette approche doit se fier à la spécification manuelle, par l'utilisateur, des véritables arêtes dans la scène. Une telle tâche peut être praticable dans le cas de scènes simples faites par l'homme où les arêtes sont peu nombreuses et facilement discernables. Par contre, l'utilité de cette méthode avec une scène naturelle et non structurée est limitée.

Une autre approche à la triangulation contrainte est décrite dans (Blanc, 1998;

Lhuillier, 1999). On utilise initialement un appariement dense entre deux images qui est ensuite converti en une triangulation de vues jointes. La triangulation est obtenue en divisant la première image en une grille de tuiles carrées. Chaque tuile de la première image se projette dans la seconde image par une homographie. Les tuiles voisines ayant des homographies suffisamment similaires sont soudées. Des essais RANSAC sont fait pour estimer la cohérence interne des appariements de chaque tuile, ce qui permet d'éliminer les appariements incorrects. Une fois que les contours polygonaux des régions appariées sont connus, les régions non-appariées sont traitées avec une triangulation de Delaunay. Cette méthode manipule correctement les régions partiellement cachées et préserve la géométrie de la scène dans les régions densément appariées. Toutefois, l'obtention d'un appariement dense initiale est nécéssaire pour produire cette triangulation. Le chapitre suivant présente une méthode de triangulation de point d'appariement épars qui utilise les images comme critère de triangulation.

Chapitre 4

Triangulation physiquement correcte: approche proposée

4.1 Les lignes directrices de l'approche proposée

Contrairement aux approches proposées dans la section 3.2.1, cette méthode est basée sur la triangulation directe des points de correspondance épars en utilisant une contrainte de triangulation basée sur les images. Les points d'appariements sont traités automatiquement afin de produire des triangles qui correspondent approximativement à des surfaces planes dans la scène. L'évaluation de la correspondance à des surfaces planes est basée sur la texture à l'intérieur des triangles. Puisque la région de support pour l'évaluation de la validité des triangles est beaucoup plus grande que celle utilisée pour l'évaluation d'un point de correspondance individuel, les résultats sont moins sensibles aux ambiguïtés locales dans les données et les erreurs de mise en correspondance dues à ces ambiguïtés sont éliminées.

La figure 4.1 présente le diagramme de flux de données de l'approche proposée.

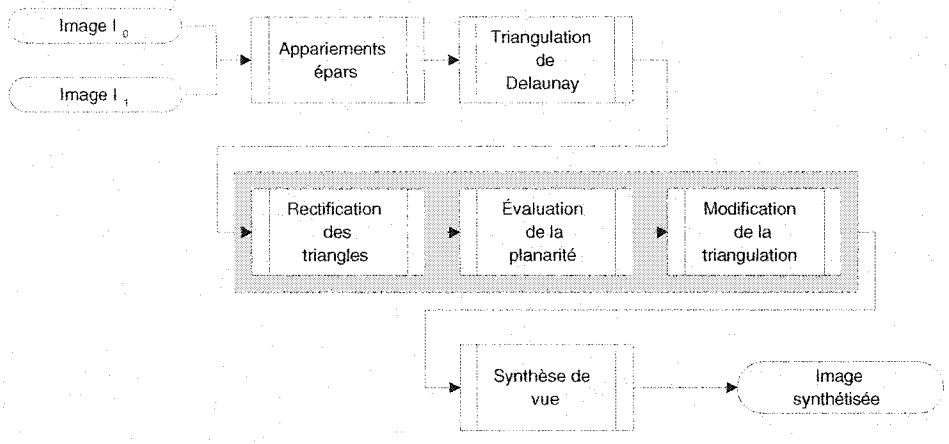


Figure 4.1 : Flux de données de l'approche proposée.

Étant donné deux vues mises en correspondance de façon éparse, une triangulation de Delaunay est utilisée comme une approximation initiale de la triangulation désirée. Cette approximation initiale est ensuite raffinée selon l'évaluation de la correspondance de chaque triangle à une surface plane dans la scène. Cette évaluation est obtenue en mesurant la corrélation entre les textures de triangles respectifs après les avoir rectifiés à un point de vue commun. Cette évaluation se base sur le fait que des triangles qui correspondent à des surfaces dans la scène auront un coefficient de corrélation élevé. La correspondance des triangles rectifiés est illustrée dans la figure 4.2. L'étape de modification de la triangulation permet de changer les triangles, mais il est possible que l'ensemble des points d'appariement initial soit déficient et empêche l'obtention d'une triangulation correcte dans certaines régions de l'image. Pour résoudre ce problème, une étape additionnelle tente d'extraire de nouveaux points d'appariement dans les régions de l'image où les triangles demeurent incorrectes. Une vérification globale pour les points ajoutés est obtenue par l'estimation de la triangulation résultante, ainsi, la susceptibilité aux erreurs d'appariements lo-

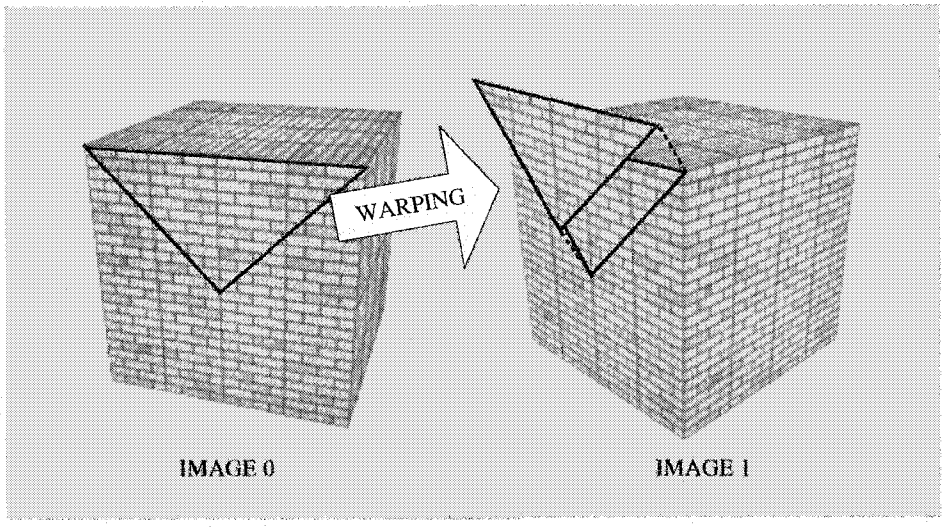


Figure 4.2 : Un triangle, créé par la connexion de trois points d'appariements dans I_0, est transformé vers son triangle correspondant dans I_1. Les textures sont ensuite comparées pour évaluer la validité physique du triangle. Une similarité élevée indique que le triangle est réellement une surface plane dans la scène.

cales est diminuée.

Cette approche de triangulation est limitée de façon inhérente par sa nature basée sur l'information contenue dans la texture. L'évaluation de la planarité peut échouer si les triangles correspondants ont des différences importantes au niveau de la taille, ou de la surface couverte dans l'image. Dans ce cas, la transformation de la texture du premier triangle vers le second va créer des artéfacts de sous-échantillonnage, particulièrement si la texture est étirée. Donc, il est possible qu'un triangle physiquement correct ne soit pas reconnu comme tel parce qu'il y a trop de différence en taille et en orientation entre les triangles correspondants dans les images sources.

Ce problème peut être limité, toutefois, en faisant la comparaison des textures dans le triangle le plus petit pour empêcher l'étirement des textures. Par contre, ceci réduit la quantité d'information utilisée et rend donc la comparaison moins fiable. La fiabilité de la comparaison est plus grande lorsque les triangles comparés sont de

tailles similaires. Donc, une mesure de fiabilité des résultats devrait tenir compte de la différence de taille des triangles comparés.

Les sections qui suivent fournissent de plus amples détails sur l'approche proposée. Le processus de rectification des triangles est décrit dans la section 4.2. L'évaluation de la présomption de planarité est discutée dans la section 4.3. Les algorithmes pour modifier la triangulation sont détaillés dans la section 4.4.

4.2 Rectification des triangles

En partant de triangles appariés dans deux vues, il est possible de vérifier leur correspondance à une surface plane dans la scène en les projetant sur un point de vue commun et en mesurant la corrélation de la texture qu'ils contiennent. Puisque le processus de rectification suppose que les triangles appariés représentent une surface plane, celui-ci va créer des distorsions si la supposition est incorrecte. La rectification des triangles peut être obtenue par un plaquage de texture affine ou perspective. Différentes approches de plaquage de texture sont discutées dans la section 4.2.1 et leur évaluation est présentée dans la section 4.2.2.

4.2.1 Déformation des triangles

Étant donné des triangles appariés dans deux vues, leur rectification à un point de vue commun peut être obtenue par plaquage de texture d'un triangle à l'autre. Ceci peut être fait avec les techniques bien connues de déformation affine ou perspective (Worlberg, 1990). La technique de déformation affine est la plus simple à calculer. Les trois sommets d'un triangle sont représentés, en coordonnées homogènes, par $A = (A_x, A_y, 1)^T$, $B = (B_x, B_y, 1)^T$ et $C = (C_x, C_y, 1)^T$. Un point $P = (P_x, P_y, 1)^T$

à l'intérieur de ce triangle peut être exprimé par la combinaison linéaire des trois sommets: $P = \alpha A + \beta B + \gamma C = M(\alpha, \beta, \gamma)^T$, où

$$M = \begin{bmatrix} A_x & B_x & C_x \\ A_y & B_y & C_y \\ 1 & 1 & 1 \end{bmatrix}$$

En utilisant les coefficients de combinaison linéaire $(\alpha, \beta, \gamma)^T$, le point correspondant P' dans le second triangle est obtenue par: $P' = M'(\alpha, \beta, \gamma)^T = M'M^{-1}P$ où M' est composée des sommets du second triangle. Ainsi, dans le cas affine, la matrice de transformation d'un point est donnée par: $H = M'M^{-1}$.

Une transformation perspective relie deux projections différentes d'une région plane. Cette relation définie une correspondance un-à-un de telle sorte que:

$$\lambda \begin{bmatrix} P'_x \\ P'_y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ 1 \end{bmatrix}$$

ou, en forme compacte, $\lambda P' = HP$ où H est une matrice 3×3 non singulière. Ce genre de transformation est appelé homographie planaire. Une homographie H n'est connue qu'a un facteur d'échelle λ près et n'a donc que 8 degrés de liberté. Ceci implique que 8 des neuf valeurs de la matrice sont linéairement indépendantes et donc, 8 équations sont nécessaires pour résoudre l'homographie. Chaque point de correspondance contribue à deux équations, donc, quatre paires de points appariés sont nécessaires pour résoudre le système d'équations linéaires.

Les transformations affine et perspective sont illustrées dans la figure 4.3. La

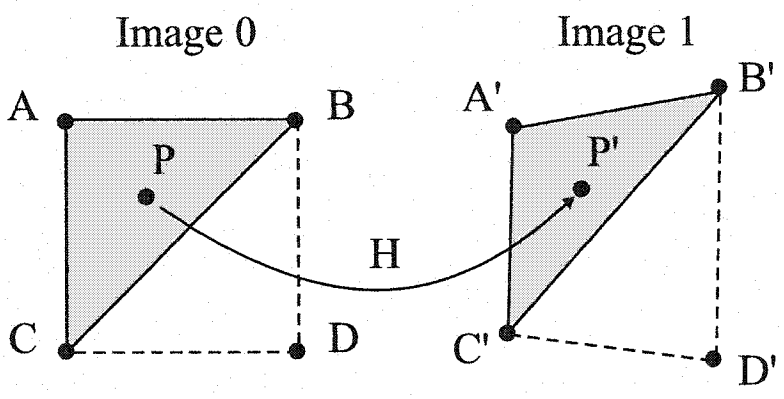


Figure 4.3 : La fonction unique qui fait correspondre un point P d'un triangle T_{ABC} vers un point P' dans un triangle $T_{A'B'C'}$ est exprimée par la matrice 3×3 H. Dans le cas d'une transformation affine, les points A,B et C seulement sont requis pour calculer H. Dans le cas d'une transformation perspective, un quatrième point, D, est nécessaire pour calculer H.

différence entre les deux méthodes de déformation réside dans le nombre de points qu'elles requièrent pour définir une transformation unique point à point. Dans le cas où les regroupements de points sont des triangles, il est plus naturel et pratique d'utiliser trois points pour spécifier une transformation.

Puisque les images sont des projections perspectives de la scène observée, la transformation perspective est plus fidèle à la réalité que la transformation affine. Étant donné qu'au moins quatre paires de points appariés sont nécessaires pour calculer une transformation perspective, il est possible de grouper les triangles par paires de voisins et d'utiliser les 4 points ainsi formés par la paire de triangles. Une autre possibilité serait d'utiliser les épipôles des images comme quatrième point. Toute homographie planaire va transformer l'épipôle de la première image sur l'épipôle de la seconde image(McMillan, 1997, Chapitre 5).

La première approche pour obtenir les quatre paires de points consiste à sélectionner une paire de triangles, en supposant qu'ils correspondent à une même surface plane dans la scène. Chaque triangle à trois voisins (à moins qu'il soit un triangle à

la bordure de la triangulation). Chaque voisin doit être considéré comme candidat possible pour former une paire. L'évaluation de la validité de cette paire de triangles adjacents peut être obtenue en transformant la paire de triangles sur leur paire correspondante dans la seconde image. Ensuite, les textures sont comparées en utilisant l'équation 4.3 définie dans la section 4.3.1. Le même processus est répété pour tous les candidats. De plus, la transformation affine est aussi utilisée avec le triangle original comme mesure de comparaison de base. L'algorithme pour former des paires de triangles est formulé comme suit:

Pour tout triangle T_i dans l'ensemble *faire*
    Pour tout voisin T_j de T_i *faire*
        Pairer T_i avec T_j et mesurer le résultat.
    Fin pour
    Garder la paire (T_i,T_j) qui maximise le résultat.
    Si résultat pairé maximum $<$ résultat affine *alors*
        Défaire la paire, T_i ne sera pas pairé.
    Fin si
Fin pour

En plus de demander plus de calcul, une seconde faiblesse de cette méthode est d'imposer le besoin de pairer les triangles ensemble à chaque fois que la triangulation est modifiée, ce qui se produit à chaque itération du processus de re-triangulation.

La seconde solution permettant d'obtenir un quatrième point suppose que la géométrie épipolaire entre les images est connue, ou a été résolue par le calcul de la matrice fondamentale. Il est possible d'obtenir la matrice fondamentale en utilisant

les points de correspondance déjà disponibles en utilisant une des méthodes décrites dans (Zhang, 1996). Néanmoins, suivant la configuration des caméras, des petites imprécisions dans les points appariés peuvent conduire à de grandes erreurs dans la position des épipôles. De plus, puisque les épipôles peuvent être loin des triangles, l'application de texture en perspective peut devenir extrêmement sensible aux instabilités numériques. En particulier, lorsque les images deviennent coplanaires, les épipôles tendent vers l'infini. Il devient donc impossible de les utiliser.

En conséquence, l'utilisation de transformation affine doit tout de même être considérée comme une approximation de la réalité pour la rectification des triangles. La comparaison des résultats de rectification obtenus par transformations affine et perspective est décrite dans la section qui suit.

4.2.2 Comparaison des transformations affine et perspective

Afin d'évaluer la validité de la transformation affine par rapport à la transformation perspective, un test contrôlé a été tenu avec des séquences d'images. Chaque séquence de test contient 6 images d'un objet tourné d'environ 10 degrés par image. En conséquence, il y a une rotation d'environ 50 degrés entre la première et la dernière image de chaque séquence. Cinq triangles, correspondant à la même surface plane de l'objet, sont définis dans la première image et suivis dans les images suivantes. La technique de comparaison des textures à l'intérieur des triangles correspondants dans chaque image de la séquence est utilisée pour comparer les deux types de transformations. Étant donné deux triangles correspondants dans deux images, cette technique consiste à transformer l'un des deux triangles afin que ses sommets coïncident à ceux de l'autre triangle. La transformation interne, affine ou perspective, est utilisée pour transférer le reste des points de la texture à l'intérieur du triangle. Ensuite, le contenu

du triangle ainsi transformé est comparé par corrélation normalisée à moyenne nulle au contenu de l'autre triangle correspondant afin de mesurer leur similarité. Pour la transformation perspective, c'est la méthode de formation de paires de triangles décrite précédemment qui a été utilisée.

La figure 5.6-a présente les 6 images de la première séquence de test. La figure 4.4 montre les triangles utilisés. Les graphiques à gauche et à droite de la figure 4.5 montrent les résultats de la corrélation des textures pour les transformations affine et perspective respectivement. Les résultats des deux méthodes de transformation sont très similaires. L'effet de la correction perspective est principalement apparent lorsque la rotation se trouve entre 30 et 40 degrés. La transformation perspective ne produit pas de résultat de corrélation beaucoup plus élevé que la transformation affine lorsque l'angle de rotation est grand. Ceci s'explique par la grande différence entre la résolution des triangles. Par exemple, le triangle #1 dans la première image de la séquence a 3395 pixels alors que le triangle correspondant dans la dernière image de la séquence n'a que 950 pixels. L'expérience précédente montre, tel que prévu, que l'influence sur les résultats de corrélation de la distorsion introduite par la transformation affine est largement inférieure à l'influence de la différence d'aire entre les triangles correspondants.

4.3 Évaluation de planarité

En se basant sur la procédure de rectification décrite dans la section précédente, on peut vérifier qu'un triangle représente une surface plane dans la scène en mesurant la corrélation entre le triangle rectifié dans une image et le triangle correspondant dans la seconde image. On s'attend à ce que la corrélation entre des triangles qui appar-

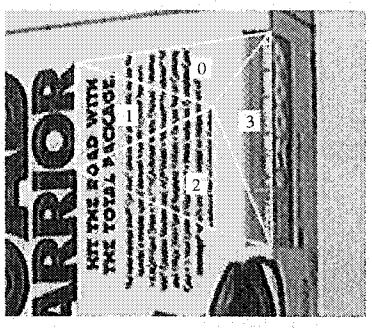


Figure 4.4 : La configuration de triangles utilisée pour le test de correction perspective. Tous les triangles sont sur la même surface plane de la boîte.

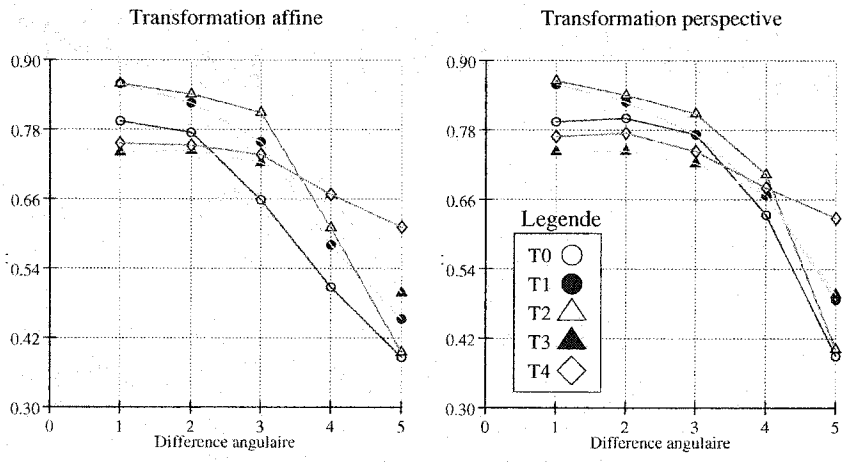


Figure 4.5 : Les graphiques à gauche et à droite montre les résultats de la corrélation de texture pour les transformations affine et perspective respectivement.

tiennent à une véritable surface plane dans la scène soit élevée alors que, autrement, le processus de rectification produira des inconsistances qui réduiront la corrélation entre les triangles correspondants. L'approche pour mesurer la similarité des triangles est discutée dans la section 4.3.1 et une évaluation détaillée de l'influence de la rotation 3D sur la mesure de similarité est présentée dans la section 4.3.2.

4.3.1 Mesure de la similarité des triangles

Étant donné deux vues, I_0 et I_1, le processus de rectification produit une troisième image I'_0, qui contient tous les triangles rectifiés qui correspondent à la déformation des triangles de I_0 vers ceux de I_1. L'évaluation de la corrélation des textures est opérée entre I'_0 et I_1 tel qu'illustré à la figure 4.6-a. La corrélation normalisée de moyenne nulle (*"zero-mean normalized cross-correlation"*, ZNCC) est utilisé à cette fin. On définit la mesure ZNCC entre deux images I_0 et I_1, dans une région carrée de $(2n+1) \times (2n+1)$ pixels centrée en (u, v), par:

$$\mathrm{ZNCC}_n(u,v) = \frac{\sum_{i=-n}^{n}\sum_{j=-n}^{n}[I_0(u+i,v+j) - \overline{I_0(u,v)}] \times [I_1(u+i,v+j) - \overline{I_1(u,v)}]}{(2n+1)^2\sqrt{\sigma_0^2(u,v) \times \sigma_1^2(u,v)}} \tag{4.1}$$

où

$$\overline{I_k(u,v)} = \frac{\sum_{i=-n}^{n}\sum_{j=-n}^{n} I_k(u+i,v+j)}{(2n+1)^2}$$

est la moyenne au point (u, v) de $I_k(k = 0, 1)$, et $\sigma(I_k)$ est l'écart type de l'image I_k dans le voisinage $(2n+1) \times (2n+1)$ de (u,v), qui est donné par:

$$\sigma(I_k) = \sqrt{\frac{\sum_{i=-n}^{n}\sum_{j=-n}^{n} I_k^2(u,v)}{(2n+1)^2} - \overline{I_k(u,v)}}.$$

L'information contenue dans des régions quasi-uniformes est très réduite, il en

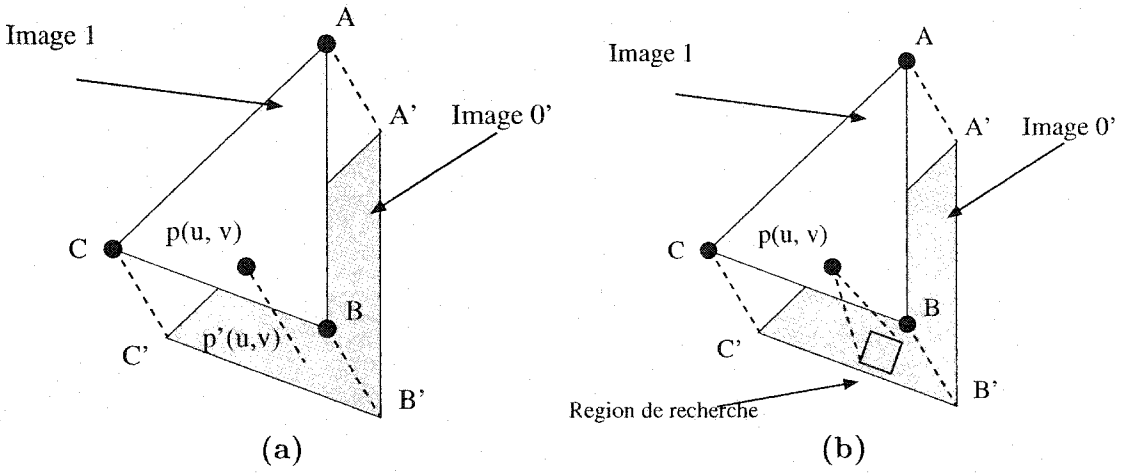


Figure 4.6 : **(a)** L'image I_0 est déformée de telle sorte que le triangle $A'B'C'$ corresponde exactement au triangle ABC dans l'image I_1. Si le triangle est physiquement correct (c'est-à-dire qu'il représente une surface plane dans la scène), alors un point $p(u, v)$ dans I_1 va correspondre au point $p'(u, v)$ à la même position dans l'image déformée I'_0. **(b)** La **région de recherche** est un carré qui entoure la position supposée du point correspondant $p'(u, v)$. Le point $p(u, v)$ sera corrélé avec tous les points dans la région de recherche. La corrélation maximale déterminera le résultat de corrélation finale pour le point $p(u, v)$.

résulte que l'évaluation de la similarité des triangles dans ces régions est peu fiable. Puisqu'il est possible que la mesure de ZNCC produise quand même des résultats élevés dans ces régions, une mesure additionnelle, appelé la variance jointe normalisée (VJN), est introduite. La VJN au point (u, v) est basée sur la variance locale dans une fenêtre carrée de $(2n+1) \times (2n+1)$ pixel autour du point et est définie comme:

$$\mathrm{VJN}(u, v) = \frac{\max(\sigma_0^2(u, v), \sigma_1^2(u, v))}{\sigma_m^2} \tag{4.2}$$

où σ_m^2 est la variance maximale possible dans les images ($\sigma_m^2 = 255^2$ pour des images avec des valeurs de tons de gris dans l'intervalle $[0, 255]$). La normalisation par σ_m^2 garantit que $\mathrm{VJN}(u, v) \in [0, 1]$. Le fait d'utiliser la variance maximale entre les deux images permet de mettre en évidence les régions différentes. En effet, si une image contient une région uniforme et l'autre une région de grande variance, on veut que la mesure de variance soit grande également afin de mettre en évidence la différence entre les images.

La mesure de VJN est utilisée pour pondérer la corrélation ZNCC afin que les régions de variance faible aient moins de poids dans l'évaluation de similarité. Pour un ensemble T_{ABC} de pixels contenus dans un triangle formé par les sommets A,B et C, la fonction de correspondance qui mesure la similarité entre les textures de ce triangle dans les images I_0' et I_1, est définie comme suit:

$$\mathrm{Match}(T_{ABC}) = \frac{\sum_{(u,v)\in T_{ABC}} \mathrm{ZNCC}_n(u, v) \cdot \mathrm{VJN}(u, v)}{\sum_{(u,v)\in T_{ABC}} \mathrm{VJN}(u, v)} \tag{4.3}$$

La fonction de correspondance est normalisée par la quantité totale d'activité (au

niveau du signal) dans le triangle. Puisque :

$$\sum_{(u,v)\in T_{ABC}} \mathrm{VJN}(u,v) \geq \sum_{(u,v)\in T_{ABC}} \mathrm{VJN}(u,v) \cdot \mathrm{ZNCC}_n(u,v)$$

on a que $\mathrm{Match}(T_{ABC}) \in [0,1]$. La fonction de correspondance, telle que définie ici, donnera une valeur élevée de corrélation dans les régions hautement texturées seulement.

Finalement, une mesure de fiabilité additionnelle de l'information (de la texture) contenue dans un triangle est définie comme la valeur moyenne de la mesure de VJN dans le triangle :

$$\mathrm{Conf}(T_{ABC}) = \frac{\sum_{(u,v)\in T_{ABC}} \mathrm{VJN}(u,v)}{\#T_{ABC}} \tag{4.4}$$

où $\#T_{ABC}$ est le cardinal de l'ensemble T_{ABC}. Cette mesure de fiabilité est utilisée dans l'étape de raffinement de la triangulation décrite dans la section 4.4.2.

La figure 4.7 résume la séquence d'opération faite durant l'évaluation de la planarité des triangles. Les vues d'origines I_0 et I_1 sont présentées dans les figures 4.7-a et 4.7-b, respectivement. L'image déformée I'_0, créée en rectifiant les triangles de I_0, est présenté dans la figure 4.7-c. Les distorsions créées par la rectification dans la texture des deux triangles incorrects sont clairement visibles. Les figures 4.7-d et 4.7-e montrent les résultats de la VJN et de la mesure ZNCC pondérée entre I'_0 et I_1 respectivement. Comme il peut être observé, la mesure ZNCC pondérée est plus basse dans les régions correspondant aux triangles incorrects.

La mesure de correspondance définie dans l'équation 4.3 est particulièrement sensible à des petites erreurs d'alignement entre les triangles. À cause des erreurs numériques et d'échantillonnage introduites par le processus de rectification, même de petites erreurs de correspondance entre les sommets des triangles peuvent systé-

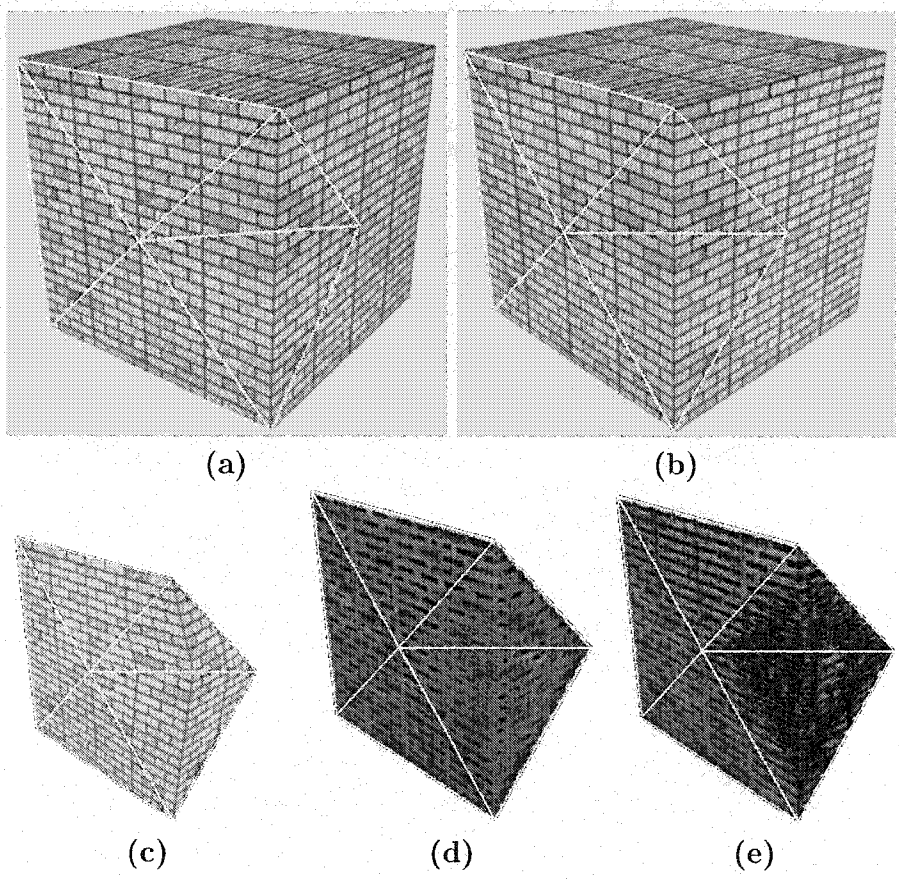


Figure 4.7 : **(a)–(b)** Les images d'origines, I_0 et I_1 respectivement, avec la triangulation de Delaunay initiales des points de correspondence. **(c)** Le résultat de la transformation des triangles de I_0 vers I_1. Cette image est dénotée par I'_0. **(d)** Le résultat de la mesure $\mathrm{VJN}(u, v)$ entre l'image rectifiée, I'_0, et l'image destination, I_1. **(e)** Le résultat de $\mathrm{VJN}(u, v) \cdot \mathrm{ZNCC}_2(u, v)$. L'effet des deux triangles incorrects est clairement visible aux endroits où l'image est plus sombre (noir indique des valeurs basses).

matiquement diminuer les résultats de corrélation. Afin de réduire l'impact de ces erreurs, l'algorithme tient compte d'une erreur possible entre la correspondance de $I'_0(u,v)$ avec $I_1(u,v)$. On suppose que le point correspondant à $I_1(u,v)$ dans l'image rectifiée, s'il existe, se trouve dans le voisinage local des coordonnées (u,v). Donc, la mesure ZNCC est évaluée entre $I_1(u,v)$ et $I'_0(u+k,v+l)$, où $|k|,|l| \leq m$. Le résultat ZNCC maximum, obtenu dans la région de recherche $(2m+1) \times (2m+1)$ autour de (u,v), remplace le facteur $\text{ZNCC}_n(u,v)$ dans l'équation 4.3. La figure 4.6-(b) donne une illustration de la région de recherche. La taille de la région de recherche a une influence sur la qualité des résultats. Cette influence sera étudiée dans la section 5.1.

4.3.2 Évaluation de l'influence de rotations 3D

Il est important d'évaluer le comportement de la mesure de similarité entre triangles définis dans l'équation 4.3 puisque cette mesure est la base du critère d'optimisation pour le processus de triangulation. Cette section examine le comportement et la robustesse de la mesure face à des transformations de la scène telles que des rotations 3D.

Afin d'évaluer la performance de la mesure face à des rotations de la scène, plusieurs tests ont été menés sur des scènes réelles et virtuelles. Dans chaque test, plusieurs images d'une scène affectée d'une rotation graduelle ont été prises. Le but du test est d'évaluer le comportement des triangles physiquement corrects par rapport à ceux qui ne le sont pas. Les images utilisées dans un des tests sont présentées dans la figure 4.8. La figure 4.8-a montre 4 des 13 images de la séquence. Les triangles utilisés dans le test sont illustrés dans la figure 4.8-b. La simplicité de la scène facilite l'observation des différents types de triangle dans la séquence. Les différents types de triangles définis dans cette scène sont :

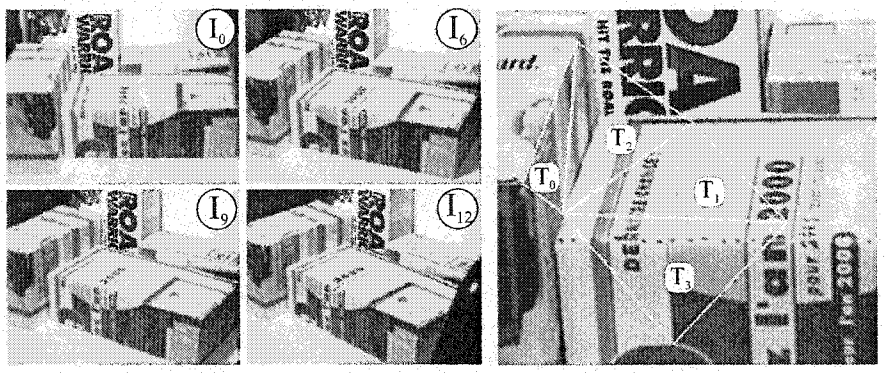


Figure 4.8 : **(a)** Images I_0, I_6, I_9 et I_{12} d'une séquence de 13 images en ton de gris de résolution 640×480 pixels. La scène est composée de boîtes texturées qui sont tournées autour d'un axe vertical commun entre chaque image. **(b)** Configuration des triangles utilisés pour le test de la mesure de similarité.

- T_0 – Triangle physiquement correct qui subit une rotation verticale.
- T_1 – Triangle physiquement correct qui subit une rotation horizontale.
- T_2 – Triangle invalide qui connecte des objets par un plan au dessus de la scène.
- T_3 – Triangle invalide qui connecte des objets par un plan au travers de la scène.

Les résultats de la mesure de similarité pour les quatre triangles étudiés dans la séquence sont montrés dans la figure 4.9. La figure 4.9-a montre la mesure de similarité obtenue sans la région de recherche. Les figures 4.9-b – 4.9-c présentent la mesure obtenue avec des régions de recherche de 3×3 et 5×5 pixels respectivement. La taille de la région de recherche permet d'augmenter la séparation relative entre les triangles valides et invalides. De plus, lorsque la région de recherche est utilisée, les pointages de correspondance des triangles sont globalement plus élevés et la séparation entre les triangles valides et invalides pour des angles de rotation très petits est plus mince. Conséquemment, la région de recherche devrait être utilisée seulement lorsque

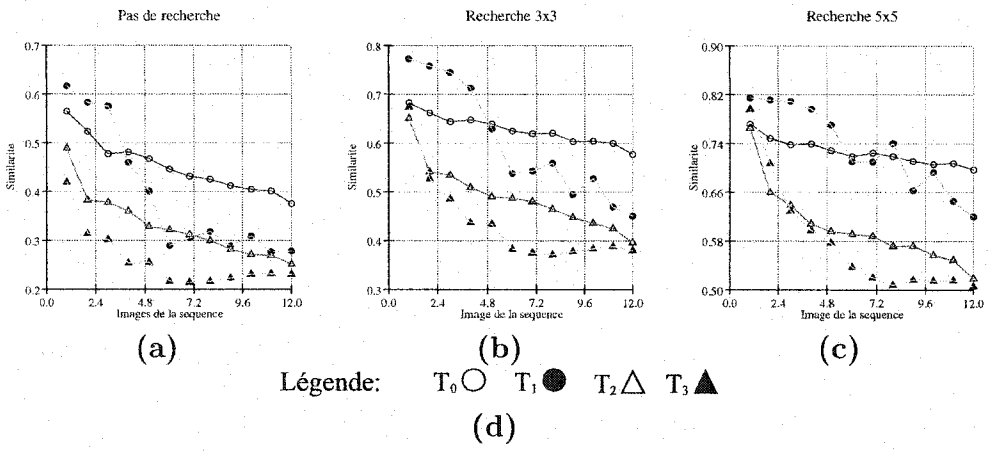


Figure 4.9 : Résultats de la mesure de similarité pour les quatre triangles montrés dans la figure 4.8-b. **(a)** Résultat de correspondance sans la région de recherche. **(b)–(c)** Résultat avec des régions de recherche de 3×3 et 5×5 pixels, respectivement. **(d)** Légende pour les triangles.

l'écart de rotation entre les vues est suffisamment élevé. Cette évaluation démontre un comportement adéquat de la mesure de correspondance. Même sans région de recherche, le pointage des triangles invalides est plus bas que les triangles valides dans la plupart des cas.

4.4 Modification de la triangulation

Lorsque la validité de tous les triangles dans les images est quantifiée, l'étape suivante de cette approche est de modifier les triangles qui ne se conforment pas à l'hypothèse de planarité. Deux étapes sont impliquées dans ce processus. Premièrement, les triangles ayant des pointages faibles sont couplés et re-triangulés par basculement des arêtes tel qu'expliqué plus bas. De façon inhérente, cette étape produite des triangles qui ont un meilleur résultat à l'évaluation de planarité et donc, qui sont plus près de la géométrie de la scène. Toutefois, il arrive que des points de correspondance man-

quant empêchent la détermination d'une triangulation correcte. Donc, une seconde étape de ce processus de triangulation se concentre sur l'ajout de point de correspondance à l'intérieur des triangles qui ont un pointage faible. La re-triangulation par basculement d'arêtes est décrite dans la section 4.4.1 et la méthode d'ajout de points d'appariements est présentée dans la section 4.4.2. Finalement, une méthode de fusion de triangles est expliquée dans la section 4.4.3.

4.4.1 Basculement d'arêtes

Étant donné la triangulation de Delaunay d'une scène et un pointage de correspondance pour chaque triangle, une méthode de re-triangulation par basculement d'arêtes est utilisée pour générer une version modifiée de la triangulation avec un pointage de correspondance amélioré. La méthode de basculement d'arêtes pour la re-triangulation est basée sur un processus itératif dans lequel les arêtes qui ne conviennent pas à une propriété prédéfinie sont sélectionnées pour basculement. Le basculement d'une arête n'est possible que si le polygone formé par les deux triangles qui partagent l'arête est convexe. On utilise l'opération de basculement d'arête pour parcourir l'espace des triangulations possibles jusqu'à atteindre la configuration qui correspond à un maximum de la fonction de pointage de la triangulation. L'opération de basculement d'arêtes est illustrée dans la figure 4.10. Dans la figure 4.10-a deux triangles voisins sont re-triangulés par basculement de leur arête commune. Dans la figure 4.10-b, les triangles voisins ne forment pas un quadrilatère convexe, le basculement de l'arête n'est donc pas possible.

Le processus de re-triangulation est utilisé pour transformer une triangulation de Delaunay en une triangulation physiquement correcte en basculant les arêtes qui ne sont pas physiquement valides. Les triangles doivent être couplés pour basculer leur

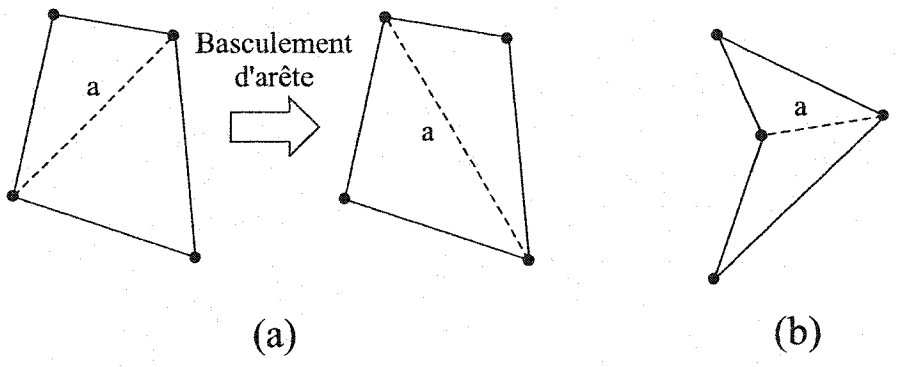


Figure 4.10 : **(a)** Basculement de l'arête a. **(b)** Le quadrilatère formé par les deux triangles est concave, par conséquent, l'arête a ne peut pas être basculé.

arête commune. La détermination de la paire est basée sur la mesure de planarité. Une paire est formée en choisissant un triangle ayant le plus bas pointage ainsi que son voisin ayant le plus bas pointage. L'arête commune est basculée, en supposant que le quadrilatère ainsi formé est convexe. Après chaque basculement, le pointage de correspondance des nouveaux triangles est évalué. Le basculement est conservé seulement si le pointage de correspondance total $\text{Match}(T_A)+\text{Match}(T_B)$ est amélioré par rapport à la configuration précédente. Un système de marquage est utilisé pour marquer les triangles pour lesquels le basculement a été testé et a échoué. Ainsi, pour former une paire, l'algorithme choisi le triangle non marqué ayant le plus bas pointage. Afin de propager la re-triangulation correctement, la marque d'un triangle est enlevée si au moins un de ses voisins a changé. Ce système de marquage garanti que l'algorithme va se terminer lorsque tous les triangles sont marqués et qu'il n'est plus possible de changer quelque chose. L'algorithme de re-triangulation proposé garantie de façon inhérente l'amélioration du pointage de correspondance totale de la triangulation modifiée. Des résultats sont montrés dans la section 5.2. L'algorithme peut être résumé par le pseudo-code suivant:

Algorithm 1 (Re-trianguler par basculement d'arêtes)

DÉBUTER

FAIRE

Mettre T_A = triangle non marqué avec le plus bas pointage.

Mettre T_B = voisin non marqué de T_A avec le plus bas pointage.

SI T_B existe ALORS

SI l'union de T_A et T_B est convexe ALORS

Générer T'_A et T'_B par basculement de l'arête commune

... de T_A et T_B.

SI $(\texttt{Match}(T'_A) + \texttt{Match}(T'_B)) > (\texttt{Match}(T_A) + \texttt{Match}(T_B))$ ALORS

Effacer la marque des voisins de T'_A et T'_B.

SINON

Récupérer l'état précédent de T_A et T_B.

Marquer T_A et T_B.

FIN SI

SINON

Marquer T_A et T_B.

FIN SI

SINON

Marquer T_A.

FIN SI

TANT QUE il y a des triangles non marqués.

FIN

Cet algorithm a été utilisé pour produire les résultats présentés dans la section 5.2.

4.4.2 Division des triangles

La re-triangulation par basculement d'arêtes produit un ensemble de triangles qui respecte la géométrie de la scène partout où il est possible de le faire. Néanmoins, certains triangles peuvent demeurer physiquement invalides parce que les points de correspondance requis ne sont pas disponibles. Donc, une méthode de raffinement est utilisée pour ajouter de nouveaux points à l'intérieur des triangles qui ont un pointage bas. Chaque point ajouté divise le triangle dans lequel il est ajouté.

La méthode de raffinement proposée commence par sélectionner un triangle qui a un pointage de correspondance bas et une mesure de fiabilité élevée (voir la section 4.3). Une mesure de fiabilité élevée indique que le triangle contient assez d'information pour être en mesure de vérifier l'exactitude d'un point ajouté. Définissons T_{ABC} comme étant le triangle sélectionné dans l'image I_0 et $T_{A'B'C'}$ son triangle correspondant dans l'image I_1. Un ensemble de coins est détecté dans l'image contenue à l'intérieur de chaque triangle. Le détecteur de coin utilisé dans les tests est décrit dans la section 2.2.3.1. Un seuil localement adaptatif est appliqué afin de garder les coins les plus proéminents dans chaque triangle correspondant. Une fois que l'ensemble de coins potentiels est extrait de T_{ABC} et $T_{A'B'C'}$, chaque paire possible de coins (D, D') est évaluée selon la fonction suivante:

$$F(D, D') = \text{Match}(T_{ABD}) + \text{Match}(T_{ACD}) + \text{Match}(T_{BCD}) \tag{4.5}$$

où T_{ABD}, T_{ACD} et T_{BCD} sont les triangles résultant de la division du triangle T_{ABC} par le point D. Les points d'appariement sélectionnés sont ceux qui maximisent la fonction $F(D, D')$ parmi l'ensemble de paire possible (D, D'). La région de support des points ajoutés est constituée du nombre de pixels inclus dans les triangles, à

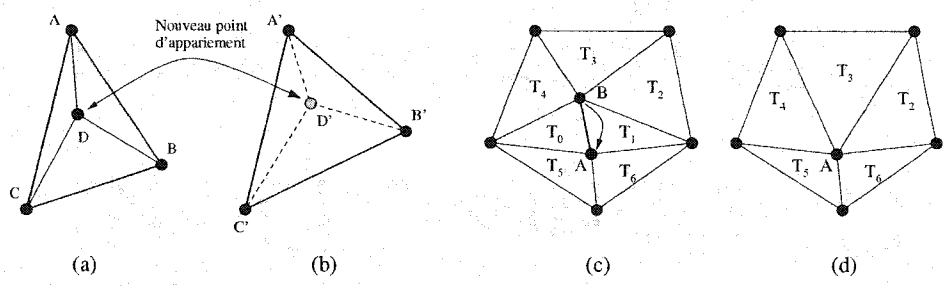


Figure 4.11 : Division et fusion de triangles. **(a)–(b)** Un triangle dans la première image est séparé en trois triangles par l'ajout d'un point D. Le point apparié D' dans le triangle correspondant est sélectionné pour maximiser le pointage de correspondance entre les sous-triangles générés. **(c)–(d)** Le sommet B de l'arête courte $\overline{AB}$ est sélectionné pour la fusion. Conséquemment, les triangles T_0 et T_1 sont enlevés en même temps que l'arête $\overline{AB}$ et le sommet B.

l'opposé d'une méthode de correspondance classique qui utilise seulement le voisinage local de chaque point. L'opération de division des triangles est illustrée dans la figure 4.11. Étant donné un point D dans un triangle de la première image, la division produit trois nouveaux triangles (figure 4.11-a). Le point apparié D' dans le triangle correspondant est choisi tel que le pointage de correspondance est maximisé pour les sous-triangles générés (figure 4.11-b).

La complexité opératoire de la recherche sur l'ensemble de toutes les paires (D, D') possibles peut être considérablement réduite par l'utilisation de la contrainte épipolaire afin d'éliminer les combinaisons de points d'appariement situés trop loin de leurs lignes épipolaires respectives. Il est possible également d'utiliser la corrélation entre les voisinages des points d'appariement afin d'éliminer davantage de combinaisons (D, D') avant d'évaluer la fonction de correspondance dans l'équation 4.5. Par contre, il faut être prudent car il est possible d'éliminer des paires de points correspondants valides à cause des distorsions perspectives qui existent entre les images avant la rectification.

Dans l'approche proposée, les points d'appariement potentiels (D, D') sont con-

traints d'être à l'intérieur des triangles originaux T_{ABC} et $T_{A'B'C'}$ afin de respecter la triangulation sous-jacente. Toutefois, il est possible que le point correspondant à un point à l'intérieur de T_{ABC} se situe à l'extérieur de $T_{A'B'C'}$ et, donc, la paire optimale de points appariés (D, D') détectée peut être incorrecte. En conséquence, des conditions supplémentaires sont utilisées pour décider si D et D' sont effectivement des points correspondants. Afin de vérifier la validité d'une paire de points appariés, ceux-ci sont ajoutés dans la triangulation en divisant les triangles auxquels ils appartiennent. Une paire est retenue si:

$$\max\{\text{Match}(T_{ABD}), \text{Match}(T_{ACD}), \text{Match}(T_{BCD})\} \geq \text{Match}(T_{ABC}) \tag{4.6}$$

ceci indique que la triangulation obtenue est améliorée. Dans un tel cas, le basculement d'arêtes est utilisé pour faciliter la réorganisation des triangles autour des points appariés. Il est toutefois possible que des paires valides ne satisfassent pas à la condition précédente puisque les triangles peuvent ne pas être physiquement corrects dans la configuration courante. Il faut donc permettre aux triangles de se réorganiser afin d'améliorer la configuration. Ainsi, une séquence de re-triangulation par basculement d'arêtes est opérée sur les nouveaux triangles. Un pointage de correspondance global est évalué et comparé à celui obtenu avant l'ajout du nouveau point d'appariement. Le point ajouté est gardé seulement si le nouveau pointage global est amélioré. Autrement, la triangulation précédente (avant l'ajout du nouveau point) est récupérée et le triangle qui n'a pas pu être divisé est marqué afin d'empêcher sa sélection dans l'itération suivante. La marque d'un triangle est enlevée lorsqu'un changement survient dans son voisinage immédiat, comme un basculement d'arête, par exemple.

Le processus complet est montré dans l'algorithme suivant :

Algorithm 2 (Re-triangulation par division de triangle)

DÉBUTER

FAIRE

Mettre T_{ABC} = triangle non marqué avec

... le plus petit pointage et la fiabilité maximale.

Enregistrer l'état courant de la triangulation.

Mettre S_{mO} = pointage globale courant.

Extraire les ensembles de points d'intérêts $\mathcal{C}(T_{ABC})$ et $\mathcal{C}(T_{A'B'C'})$.

Selectionner $(D*, D*')$ tel que $F(D, D')$ est maximal sur

... $\mathcal{C}(T_{ABC}) \times \mathcal{C}(T_{A'B'C'})$

Utiliser $(D*, D*')$ pour diviser les triangles T_{ABC} et $T_{A'B'C'}$.

SI $\max\{\text{Match}(T_{ABD*}), \text{Match}(T_{ACD*}), \text{Match}(T_{BCD*})\} \geq \text{Match}(T_{ABC})$ ALORS

Ajouter la paire $(D*, D*')$ et garder la division.

Enlever la marque sur T_{ABD*}, T_{ACD*}, T_{BCD*}, et leurs voisins.

SINON

Re-triangulater par basculement d'arêtes (Algorithme 1).

Mettre S_{mN} = pointage globale courant.

SI $S_{mN} > S_{mO}$ ALORS

Ajouter la paire $(D*, D*')$ et garder la division.

Enlever la marque sur T_{ABD*}, T_{ACD*}, T_{BCD*}, et leurs voisins.

SINON

Rejeter la paire $(D*, D*')$.

Rétablir l'état préalable de la triangulation.

Marquer le triangle T_{ABC}.

FIN SI

FIN SI

TANT QUE il y a des triangles non marqués et

... Match(T_{ABC}) est suffisamment bas.

FIN

Le pointage global S_m, mentionné dans la description ci-dessus, est défini comme le pointage de similarité moyen par triangle. De la même façon, un pointage global pondéré S_w est défini en utilisant l'aire des triangles pour déterminer leur poids individuel dans le pointage global. La définition de ces pointages de similarité est donnée par:

$$S_m = \frac{\sum_{T_i \in \mathcal{T}} \text{Match}(T_i)}{\#\mathcal{T}} \quad ; \quad S_w = \frac{\sum_{T_i \in \mathcal{T}} \text{Match}(T_i) \cdot \text{Area}(T_i)}{\sum_{T_i \in \mathcal{T}} \text{Area}(T_i)} \tag{4.7}$$

où $\mathcal{T}$ est l'ensemble de triangles dans la scène. Le pointage global pondéré est utilisé dans une étape subséquente de l'approche proposée.

4.4.3 Fusion de triangles

Le processus de division, décrit dans la section précédente, est utilisé pour ajouter de nouveaux points de correspondance afin d'améliorer et raffiner la triangulation. Toutefois, les points de correspondance ajoutés, ou même les points initiaux, peuvent être incorrects. Afin de gérer cette éventualité, l'approche proposée applique une étape additionnelle de fusion de triangles dans laquelle certains points de correspondance peuvent être enlevés. Afin de minimiser l'impact sur le pointage global, une méthode de suppression de points sélectionne des triangles qui ont une ou des arêtes courtes et

élimine l'arête en enlevant un des sommets. Le processus de fusion de triangles par élimination d'arêtes est illustré dans la figure 4.11. Dans la figure 4.11-c, le sommet B d'une arête courte $\overline{AB}$ est sélectionné pour être éliminé. En conséquence, dans la figure 4.11-d, les triangles T_0 et T_1 sont supprimés en même temps que l'arête $\overline{AB}$ et le sommet B.

L'élimination de sommets est un processus itératif qui utilise, tout comme les processus précédents, une méthode de marquage afin de propager les opérations d'une itération à l'autre. À chaque itération, une arête non marquée avec le pointage le plus bas est sélectionnée comme candidat pour l'effondrement. Le pointage d'une arête est défini par la longueur de cette arête modulée par la moyenne des pointages de similarité des triangles qui détiennent cette arête. Ainsi, les arêtes courtes des triangles avec de mauvaises mesures de similarité seront choisies comme candidats pour l'élimination. Afin de décider si l'arête sélectionnée devrait être éliminée et pour identifier quel sommet de l'arête sera supprimé, le pointage global pondéré S_w, défini dans l'équation (4.7), est utilisé. La suppression de chaque sommet de l'arête sélectionnée est effectuée et le pointage global pondéré est mesuré dans les deux cas. L'élimination qui a produit le pointage global le plus élevé est gardée, en supposant que le pointage global est amélioré par rapport au pointage original sans l'élimination de l'arête. Si l'arête n'est pas supprimée, elle est marquée afin d'éviter qu'elle soit choisie dans les itérations suivantes.

Les sommets qui font partie de l'enveloppe convexe de la triangulation ne sont jamais supprimés afin d'éviter la réduction de la surface de la scène couverte par la triangulation. Les sommets connectés à un ensemble de triangles qui forment un polygone non convexe ne peuvent pas être supprimés non plus. Lorsque les triangles forment un ensemble convexe, l'élimination du sommet provoque des replis dans la

triangulation où les triangles se superposent. L'algorithme de fusion de triangles est résumé dans le pseudocode suivant :

Algorithm 3 (Re-trianguler par fusion de triangle)

DÉBUTER

Marquer les arêtes appartenant à l'enveloppe convexe

... de la triangulation.

FAIRE

Sélectionner $\overline{AB}$ = arête non marqué ayant le pointage minimal.

SI les triangles connectés à $\overline{AB}$ forment un ensemble convexe ALORS

Mettre S_{wO} = pointage global pondéré courant.

Calculer S_{wA} = pointage global pondéré lorsque le sommet A

... est supprimé.

Calculer S_{wB} = pointage global pondéré lorsque le sommet B

... est supprimé.

SI $\max\{S_{wA}, S_{wB}\} < S_{wO}$ ALORS

Rejeter l'effondrement et marquer l'arête $\overline{AB}$.

SINON

SI $S_{wA} > S_{wB}$ ALORS

Supprimer le sommet A.

ELSE

Supprimer le sommet B.

FIN SI

Enlever la marque des triangles modifiés et leurs voisins.

FIN SI

SINON

Marquer l'arête $\overline{AB}$.

FIN SI

TANT QUE il y a des arêtes non marquées et que le pointage des arêtes ... sélectionnées est suffisamment bas.

FIN

4.4.4 L'algorithme complet

Les opérations itératives, décrites dans les sections 4.4.1 à 4.4.3, de basculement, division et fusion sont combinées dans l'approche proposée afin de modifier la triangulation initiale pour améliorer le pointage de similarité. Leur application combinée est composée d'une séquence d'opérations de basculement, division, basculement et fusion. Les opérations de basculement permettent de réorganiser les triangles autour de chaque changement, que ce soit une division ou une fusion. Ainsi, l'effet hérité de la triangulation incorrecte est graduellement éliminé. Il est à noter que, dans la phase de division, les points d'appariements erronés génèrent des triangles de petite taille dans leur voisinage à cause de la distorsion qu'ils introduisent. Ces petits triangles sont ensuite fusionnés dans la phase de fusion et les points erronés sont supprimés. Néanmoins, selon l'erreur dans les points d'appariements initiaux, plusieurs itérations de division, basculement et fusion peuvent être requises. L'algorithme complet pour la re-triangulation est résumé dans le pseudocode suivant, où il est à noter que l'algorithme 2 contient déjà une séquence de divisions et basculements:

Algorithm 4 (Re-triangulation)

DÉBUTER

```
Créer une triangulation initiale par la méthode de Delaunay.
Enlever toutes les marques des triangles et des arêtes.
Re-trianguler par basculement d'arêtes (algorithme 1).
FAIRE
    Re-trianguler par division de triangles (algorithme 2).
    Re-trianguler par fusion de triangles (algorithme 3).
TANT QUE le pointage globale continue de croître.
FIN
```

La boucle principale de l'algorithme 4 peut continuer très longtemps avant d'arrêter. Puisqu'il est impossible d'atteindre un pointage parfait, il y aura toujours des modifications à faire à la triangulation. Une application de cet algorithme peut définir un pointage minimal (< 1) à atteindre et arrêter la boucle principale. De plus, une application qui a une contrainte de temps pourrait également donner un nombre d'itérations maximal pour l'algorithme.

Chapitre 5

Resultats

Ce chapitre discute des résultats des méthodes et algorithmes décrits jusqu'à maintenant. La première série de résultats, présentée dans la section 5.1, concerne la mesure de planarité des triangles basée sur l'image qu'ils contiennent. Ensuite, l'évaluation de planarité est utilisée pour l'algorithme de basculement d'arêtes. Des résultats de retriangulation utilisant cet algorithme sont montrés et discutés dans la section 5.2. Des résultats de raffinement de triangulation accompagné d'exemples de synthèse de vue sont montrés dans les sections 5.3, 5.4 et 5.5.

5.1 Résultats de l'évaluation de planarité

L'évaluation de planarité est la mesure de base utilisée dans les autres algorithmes. Il faut vérifier son comportement afin d'être certain que cette mesure satisfait réellement aux besoins des algorithmes de triangulation. Afin de tester la mesure, des séquences d'images de différentes scènes sont prises. Ensuite, des triangles prédéterminés sont créés dans des situations contrôlées afin d'observer le comportement de la mesure dans chaque situation.

5.1.1 Résultats de la mesure sur un cube en rotation

Ce test montre comment la mesure est affectée par des rotations en 3 dimensions. La rotation en profondeur (autour de l'axe Y) est particulièrement intéressante à observer puisqu'elle produit des distorsions perspectives entre les images. Ce test utilise 10 images d'un cube, chaque image est prise d'un point de vue tourné d'environ 10 degrés autour de l'axe Y (axe vertical) du cube. La figure 5.1-(c) illustre la séquence. De plus, la séquence a été produite avec deux textures différentes afin d'évaluer l'influence de la texture sur le comportement de la mesure. Ces textures sont montrées dans les figures 5.1-(a) et (b). La texture 0 est un marbre de Perlin et la texture 1 est un patron de briques grises. La texture de marbre de Perlin contient beaucoup d'information et ne se répète pas. À l'opposée, la texture de brique contient moins d'information et se répète le long de la surface visible. Des images synthétiques ont été utilisées pour ce test afin de mieux contrôler les paramètres de point de vue et de projection perspective.

Afin de tester la mesure de planarité, deux ensembles de triangles ont été créés. Le premier groupe est composé de 4 triangles physiquement corrects montrés dans la figure 5.2-(a). Le second groupe, montré dans la figure 5.2-(b), contient 5 triangles. Deux de ces triangles, T_2 et T_4, sont physiquement incorrects. Le test consiste à prendre chaque image de la séquence et le comparer avec l'image 0. Les points d'appariements sont spécifiés manuellement, ce qui veut dire qu'une erreur d'environ 1,5 pixel autour de chaque point est possible. C'est pourquoi un cube a été utilisé, afin de simplifier autant que possible la procédure de mise en correspondance.

Les résultats des mesures pour les triangles physiquement corrects sont affichés dans la figure 5.3-(a) pour la texture 0 et dans la figure 5.3-(b) pour la texture 1. Puisque la forme de la scène et la configuration des triangles sont identiques entre les

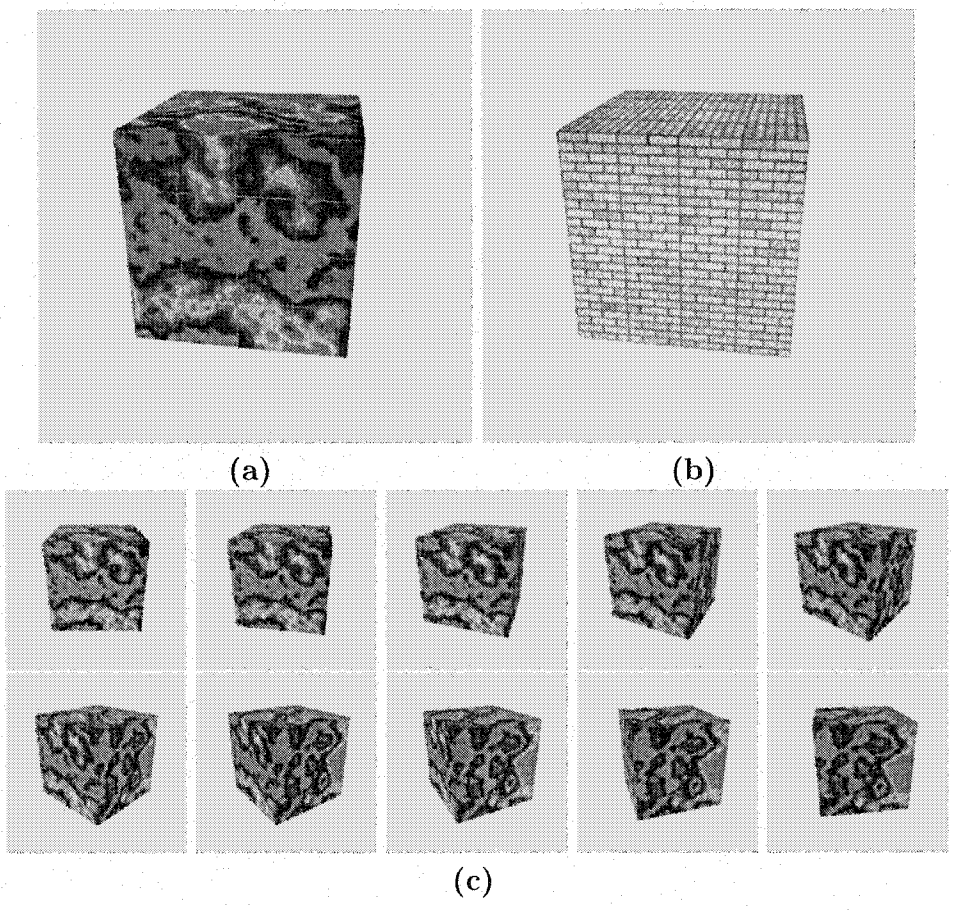
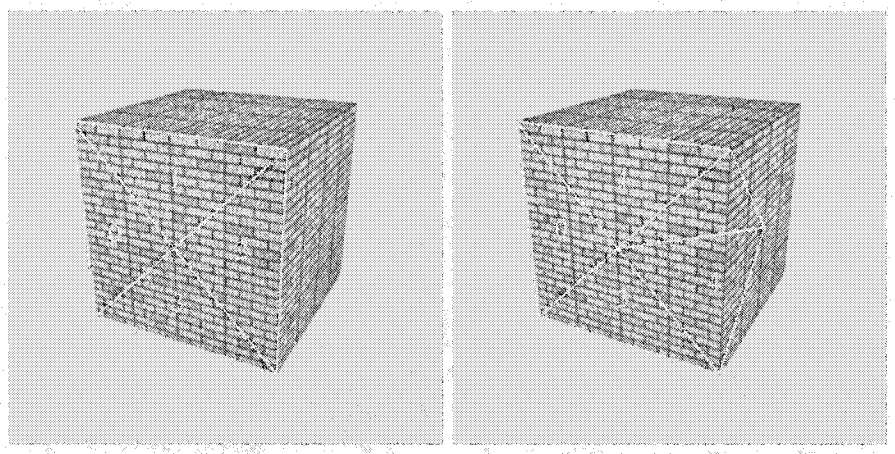


Figure 5.1 : **(a)-(b)** Textures utilisées pour les séquence 0 et 1 respectivement. **(c)** Images 0 à 9 de la séquence de cube en rotation avec la texture 0.



(a) (b)

Figure 5.2 : **(a)** 4 triangles utilisés pour le test du comportement de la mesure avec des triangles physiquement corrects. **(b)** 5 triangles pour le test des triangles incorrects. Les triangles (2) et (4) sont physiquement incorrects.

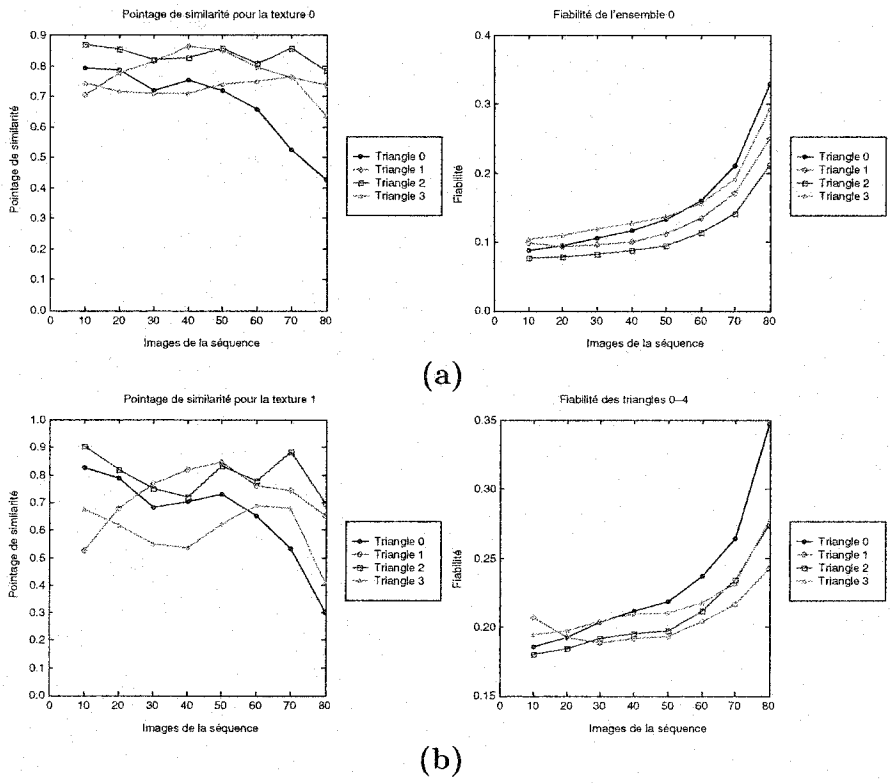


Figure 5.3 : Résultats pour la similarité et la fiabilité pour les images 1 à 8 comparées à l'image 0. **(a)** Résultats avec la texture 0. **(b)** Résultats avec la texture 1.

deux tests, il est attendu que la forme de la courbe soit la même malgré la différence de texture. C'est effectivement ce que l'on observe au niveau de la mesure de similarité. La mesure de variance (fiabilité) est croissante parce que les triangles deviennent de plus en plus petits en s'approchant de l'image 8. Donc, la rectification amène un triangle plus grand dans un triangle plus petit et donc, condense l'information de texture, ce qui augmente la variance moyenne par pixel. Le but du pointage de fiabilité est de mesurer le niveau de confiance de la mesure de similarité. Lorsque beaucoup d'information de texture est présente dans l'image, on présume que la fiabilité est bonne. Toutefois, lorsque la scène est tournée de façon à ce que la texture soit projetée dans des triangles plus petits, l'information de texture moyenne par pixel augmente, mais ceci ne veut pas dire que la fiabilité est meilleure. Afin de compenser cet effet secondaire du changement de résolution des triangles correspondant d'une image à l'autre, il faudrait également utiliser le rapport des aires des triangles (aire du triangle source par rapport à l'aire du triangle destination). Lorsque ce rapport est plus grand que 1, le triangle source est plus grand que le triangle destination et donc, la texture subit une compression, ce qui implique que la variance moyenne augmente, il faut donc diviser la variance moyenne par le rapport des aires, neutralisant ainsi l'effet trompeur du changement de résolution.

On peut également observer que le pointage de similarité décroît lorsque la rotation entre les images comparées augmente. Par exemple, la différence angulaire entre les images 0 et 8 est de 80 degrés. Dans ce cas, la rectification transforme les triangles qui sont vus de face dans l'image 0 vers des triangles qui sont vus presque complètement de côté dans l'image 8. Donc, il est normal que le pointage de similarité baisse dans ce cas, c'est-à-dire lorsque l'aire visible des triangles comparés est très différente. Ceci pourrait également être pris en compte dans la mesure de fiabilité.

Les résultats du test avec 3 triangles corrects et 2 triangles incorrects sont montrés dans les figures 5.4 et 5.5. Les triangles T_0, T_1 et T_3 sont physiquement corrects alors que les triangles T_2 et T_4 ne le sont pas. Pour ce test, on compare les images 3, 4, 5, 6, 7 et 8 avec l'image 2 puisque les deux faces du cube doivent être visible. Les graphiques 5.4-(a) et 5.5-(a) montrent le résultat de la mesure sans l'usage de la région de recherche. Les figures 5.4-(b) et 5.5-(b) montrent les résultats avec l'utilisation d'une région de recherche 3×3, telle que décrite dans la section 4.3.1. Les graphiques montrent que, lorsque la fiabilité est relativement bonne, pour les comparaisons 1 à 5, les triangles incorrects ont toujours un pointage de similarité plus bas que les triangles corrects. L'utilisation de la région de recherche accroît la différence de pointage entre les triangles corrects et incorrects. Ceci peut sembler avantageux à première vue, toutefois, pourvu que les triangles incorrects aient un pointage plus bas que les triangles valides, augmenter la distance entre les pointages ne change pas vraiment le comportement des algorithmes. Par contre, la région de recherche devient beaucoup plus utile quand la rotation des points de vue entre les images comparées devient plus importante. Par exemple, lorsque l'on compare les images 2 et 8 sans région de recherche (dernière comparaison sur les graphiques), certains triangles corrects reçoivent un pointage plus bas que les triangles incorrects. En utilisant la région de recherche, la situation est corrigée et les triangles incorrects reçoivent toujours un pointage inférieur.

Dans la figure 5.5-(b), la texture utilisée est une texture de briquesqui contient moins d'information que la texture 0 (marbre de Perlin). La texture influence quelque peut le comportement de la mesure de similarité lorsque la transformation est trop grande. C'est le cas dans la figure 5.5-(b) où un triangle physiquement correct a un pointage inférieur à un triangle physiquement incorrect. Toutefois, la rotation est de

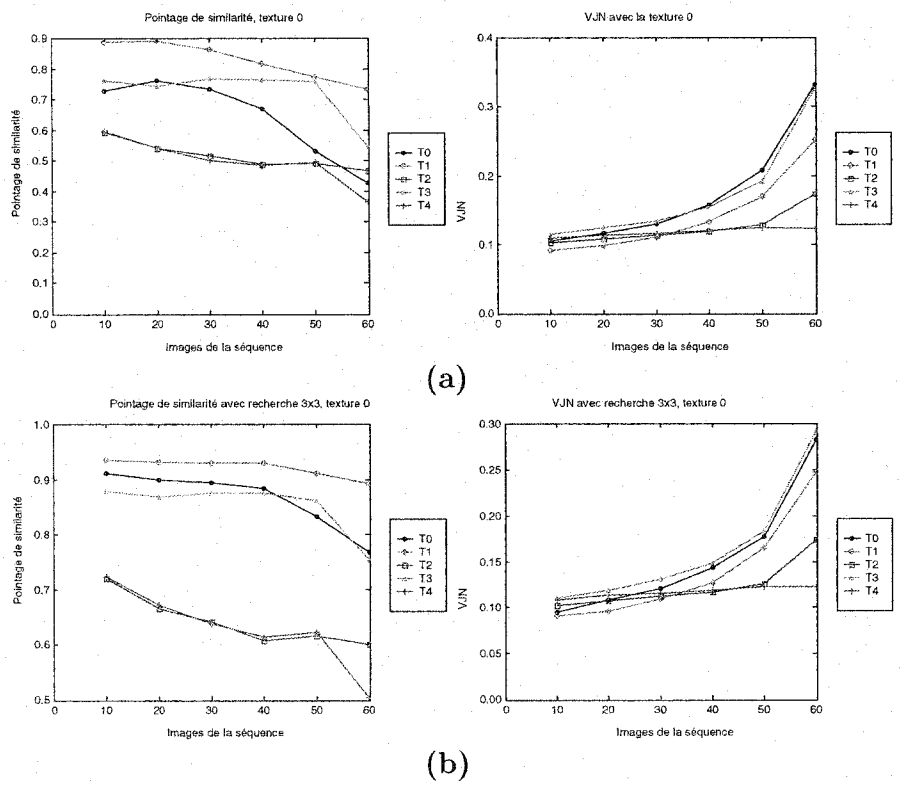


Figure 5.4 : Résultats de la mesure entre les images 3 à 8 avec l'image 2 en utilisant la texture 0. **(a)** Pas de région de recherche. **(b)** Région de recherche de 3×3.

60^o ce qui est extrême et montre la limite pratique de la méthode de comparaison des triangles.

5.1.2 Résultats de la mesure sur une boîte en rotation (scène réelle)

Le test précédent a été fait sur une séquence d'images générées par un logiciel de CAO. Ce test a permis de tester la mesure de planarité avec des conditions d'éclairages parfaitement contrôlées, aucun bruit dans les images. De plus, les positions de caméra ainsi que la projection perspective sont parfaitement connues. Puisque la méthode sera utilisée pour des images réelles, l'étape suivante est de faire un test similaire

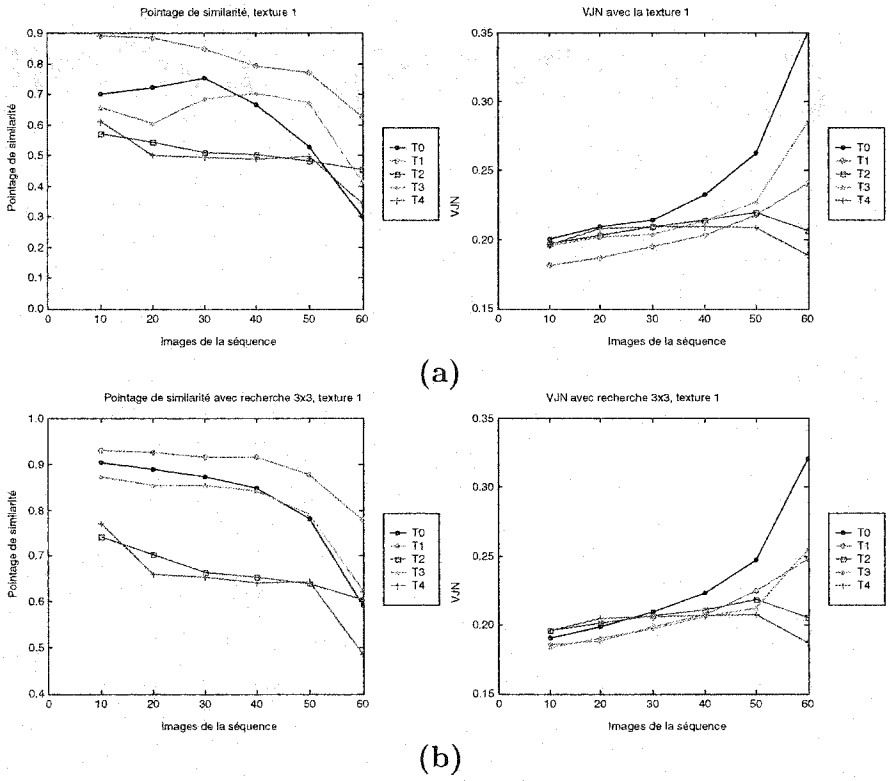
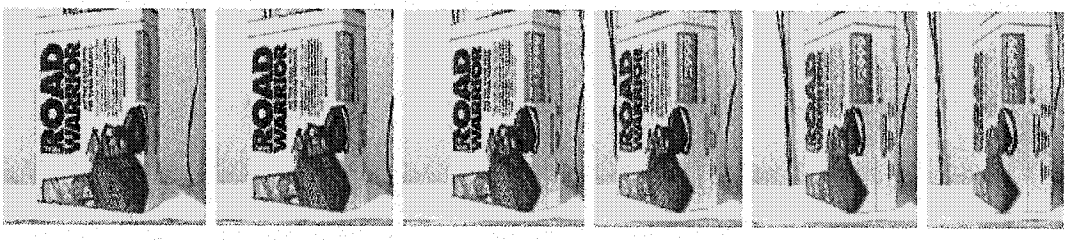
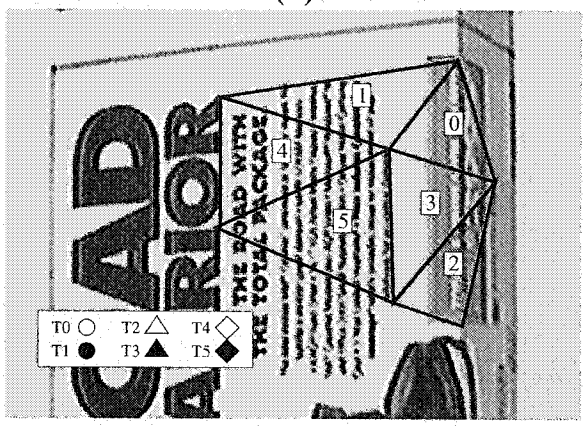


Figure 5.5 : Résultats de la mesure entre les images 3 à 8 avec l'image 2 en utilisant la texture 1. **(a)** Pas de région de recherche. **(b)** Région de recherche de 3×3.



(a)



(b)

Figure 5.6 : **(a)** 6 images de la séquence de la boîte en rotation (images numéro 30,40, 50, 60, 70 et 80). **(b)** Triangles utilisés dans le test. Les triangles T_0, T_2 et T_3 sont considérés comme incorrects puisqu'ils couvrent deux faces de la boîtes. Les triangles T_1, T_4 et T_5 sont considérés comme corrects parce qu'ils sont sur la même surface de la boîte.

avec des images réelles. La scène qui sera utilisée pour le test est montrée dans la séquence d'images de la figure 5.6-(a). On y montre une boîte en rotation autour d'un axe vertical. La boîte subit une rotation d'environ 10 degrés entre chaque image. Les images sont numérotées par leur angle de rotation autour de l'axe vertical. Donc, l'image 30 montre la boîte tournée de 30° par rapport à la position frontale. Les images ont une résolution de 640x480 pixels et on un niveau de bruit non-négligeable. La figure 5.6-(b) montre la configuration des triangles utilisée pour le test. L'ensemble est composé de trois triangles physiquement corrects, numérotés comme T_1, T_4 et T_5. Ces triangles correspondent à une surface plane dans la scène, dans ce cas, la face frontale de la boîte. Les autres triangles, T_0, T_2 et T_3, sont considérés comme physiquement incorrects parce qu'ils ne correspondent pas à une vraie surface plane dans la scène. Les points d'appariements entre chaque image ont été choisis manuellement autour de points clairement visibles pour un maximum de précision. Mais, il reste toutefois une erreur d'environs 1 pixel sur la position de ces points. Les triangles sont également spécifiés manuellement.

Les résultats sont montrés dans la figure 5.7. L'axe horizontal des graphiques représente la différence angulaire entre les images qui sont comparées. Les images 40 à 80 ont été comparées à l'image 30. Donc, la première graduation sur le graphique est de 10° qui correspondent à la comparaison de l'image 40 avec l'image 30. Ensuite, 20° pour la comparaison de l'image 50 avec l'image 30, et ainsi de suite. Ce test permet d'observer l'influence de rotation en 3D sur la mesure. Un autre aspect intéressant à observer est l'influence de la région de recherche sur la mesure. On cherche à voir si l'utilisation d'une région de recherche améliore la capacité de la mesure à différencier les triangles physiquement corrects de ceux qui ne le sont pas. Afin de vérifier ceci, la séquence a été testée trois fois. La figure 5.7-(a) montre les résultats sur la séquence

sans utiliser de région de recherche. Les figures 5.7-(b) et 5.7-(c) montrent les résultats de la mesure avec une région de recherche de 3×3 et 5×5 respectivement.

On peut observer que la mesure de similarité a tendance à décroître significativement, même pour les triangles physiquement corrects, lorsque l'angle de rotation est plus grand que 30°. Cette observation peut s'expliquer par les faits suivants :

- Des distorsions dans les images rectifiées sont introduites par la transformation affine. Puisque la calibration et la géométrie 3D de la scène sont inconnues, le processus de rectification est simplement une transformation affine 2D. Ainsi, lorsque la différence angulaire entre les triangles augmente, l'erreur entre la transformation affine et la vraie transformation perspective devient plus grande.

- La différence entre la taille des triangles correspondants introduit une différence croissante entre les images à cause des différentes résolutions d'échantillonnage de chaque image. Par exemple, dans l'image 80, le triangle T_5 prend beaucoup moins d'espace que dans l'image 30. Donc, lorsque les deux triangles sont comparés, même s'ils sont rectifiés dans un plan commun, une grande partie d'erreur d'appariement est introduite à cause de la différence d'échantillonnage. Le processus de rectification ne peut pas compenser pour l'information qui n'existe pas dans le triangle le plus petit. La rectification actuelle ne peut pas compenser pour une telle différence de taille. Il faudrait utiliser un mécanisme de filtrage de texture plus complexe, comme les mipmaps ou du filtrage anisotropique pour préserver le contenu fréquenciel dans l'image lorsque les triangles sont rectifiés.

On peut également observer l'effet de la taille de la région de recherche. La région de recherche a tendance à amener les pointages de similarité plus près les uns des autres. La distance minimum entre les groupes de triangles corrects et incorrects

avec une différence de 10 degrés est :

- 0.196 sans région de recherche;
- 0.147 avec une région de recherche de 3×3;
- 0.066 avec une région de recherche de 5×5.

Une région de recherche plus grande tend à réduire la distance absolue dans l'espace des pointages de similarité entre tous les triangles. Mais, si on mesure la distance relative entre les groupes de triangles corrects et incorrects par rapport à la variance de chaque groupe, on voit que la distance entre les groupes est plus grande avec une région de recherche de 3×3 que sans région de recherche. L'utilisation d'une région de recherche de 5×5 pixels est également bonne en terme de regroupement des pointages, mais elle augmente significativement le temps d'exécution. La région de recherche de 3×3 semble être un bon compromis entre un bon regroupement et le temps d'exécution requis.

La mesure de planarité utilise la corrélation normalisé de moyenne nulle, tel que définit dans l'équation 2.11, afin d'évaluer la correspondance entre l'image rectifiée et l'image destination. Toutefois, pour évaluer l'influence des autres types de corrélation, la scène de la figure 5.6, avec les mêmes triangles, a été utilisée pour conduire des tests utilisant d'autres mesures de corrélation. La figure 5.8-(a) illustre la mesure de planarité utilisant la corrélation normalisée telle que définit dans l'équation 2.11. La figure 5.8-(b) montre les résultats de la mesure de distance euclidienne tel que définit dans l'équation 2.12. Il est clair que le comportement de la mesure de planarité est meilleur en utilisant la mesure ZNCC. Il était à prévoir qu'il en soit ainsi puisque la mesure ZNCC est beaucoup plus robuste aux changements d'illumination et aux différences d'échelle globales entre les images.

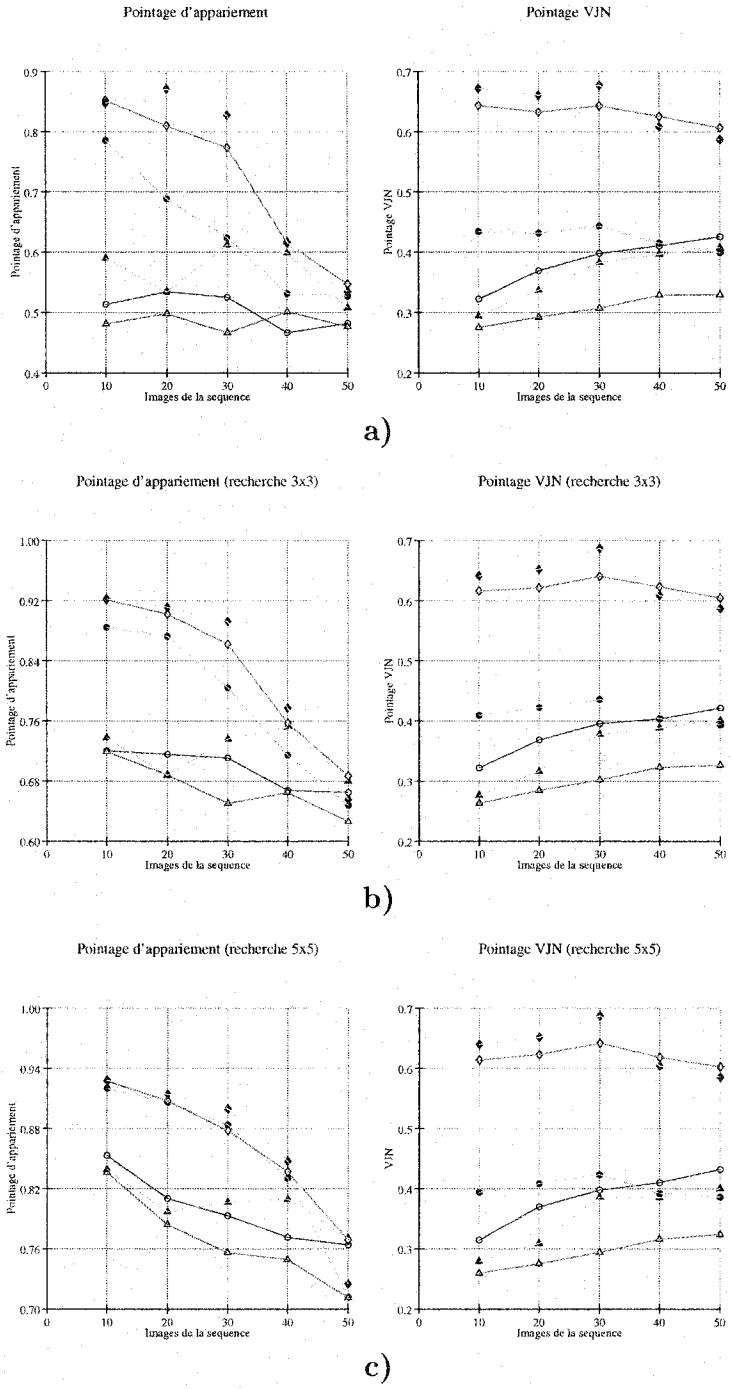


Figure 5.7 : Graphiques montrant la mesure entre les images 40 à 80 et l'image 30 de la scène de la boîte. **(a)** Sans région de recherche. **(b)** Avec une région de recherche de 3×3. **(c)** Avec une région de recherche de 5×5.

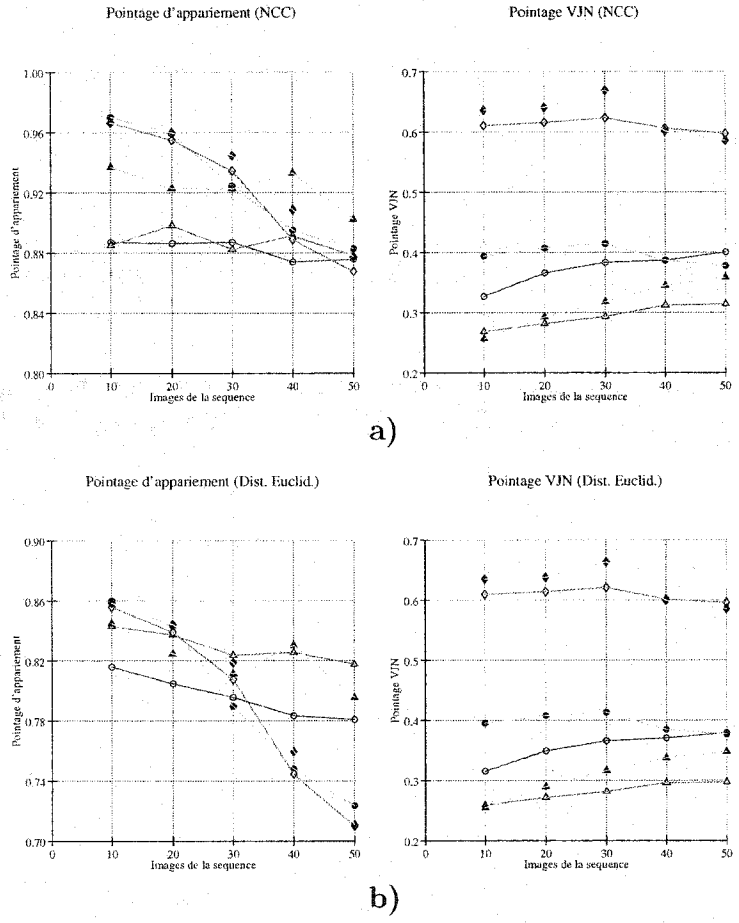


Figure 5.8 : Résultat de la mesure entre les images 40 à 80 avec l'image 30 de la scène de la boîte. La mesure utilise une région de recherche de 3×3. **(a)** NCC (Corrélation normalisée). **(b)** Distance euclidienne.

5.2 Résultat de retriangulation par basculement d'arêtes

Cette section présente quelques résultats de l'algorithme de re-triangulation. Les exemples montrés sont composés de scènes simples afin de démontrer clairement l'effet de la triangulation. La première scène, montrée dans les figures 5.31-a et 5.31-b, est constituée d'une pile de boîtes de différentes textures où les points d'appariement sont situés sur les coins visible des boîtes. Les résultats de la transformation d'une vue sur l'autre basée sur une triangulation de Delaunay initiale des points d'appariements est montrée dans la figure 5.9-a. Les divergences générées par les triangles invalides sont visibles sur chacune des trois grandes boîtes. La figure 5.9-b présente la transformation obtenue en utilisant la triangulation produite par l'algorithme proposé. Comme on peut le voir, les triangles qui étaient invalides ont été corrigés. De plus, certains triangles qui étaient déjà valides sur les petites boîtes ont été changés pour correspondre à une meilleure configuration. Le changement des triangles valides se produit parce que l'algorithme de re-triangulation cherche à augmenter le pointage de correspondance. Bien que moins visibles, ces changements dans la triangulation réduisent la distorsion introduite par la transformation affine appliquée sur les textures des triangles. Les améliorations quantitatives sur les pointages de similarités sont présentées dans les graphiques de la figure 5.9. Dans ces graphiques, chaque point représente le pointage de similarité (axe horizontal) et le pointage de fiabilité (axe vertical). Les figures 5.9-c et 5.9-d montrent, respectivement, la distribution des pointages de similarité avant et après la re-triangulation. Le pointage de similarité des triangles individuel est amélioré, tel que désiré.

Les résultats pour une seconde scène, composée d'images réelles de la séquence

illustrée dans la figure 4.8, sont montrés dans la figure 5.10-a. On peut y voir plusieurs triangles incorrects qui sont tous corrigés dans la figure 5.10-b. De plus, on peut remarquer l'amélioration du pointage dans les figures 5.10-c et 5.10-d. Plusieurs triangles avaient un pointage en dessous de 0.46, mais après la re-triangulation, ces triangles ont cessé d'exité pour être transformé en nouveaux triangles avec un meilleur pointage.

Les images dans les deux exemples contiennent des arêtes fortes directement sur les surfaces planes des boîtes, mais il n'y a pas de telles arêtes pour définir clairement la limite entre chaque surface plane. Donc, une triangulation contrainte basée sur l'extraction des arêtes telle que décrite dans (Havaldar et al., 1996; Qian Chen, 1997) ne fonctionnerait pas alors que la méthode proposée produit une triangulation correcte.

5.2.1 Basculement d'arêtes sur une scène complexe

Puisque les exemples précédents étaient des scènes simples, il est difficile d'évaluer le comportement de l'algorithme pour des scènes plus complexes. D'autres tests ont été faits en utilisant des images de scènes plus complexes pour voir si l'algorithme pouvait toujours améliorer la triangulation.

La scène du laboratoire, illustrée dans la figure 5.11, est considérée comme complexe. Cette scène est composée de deux images. Les images ne sont pas calibrées, ce qui signifie que la rectification était une transformation affine seulement et les points de correspondance ont été obtenus manuellement sans méthode spécifique. Il y a beaucoup d'occlusions dans la scène et les points d'appariement ne sont pas souvent sur les coins des objets. Donc, il ne sera pas possible pour la méthode de basculement d'arêtes d'obtenir une triangulation valide pour tous les triangles. Les

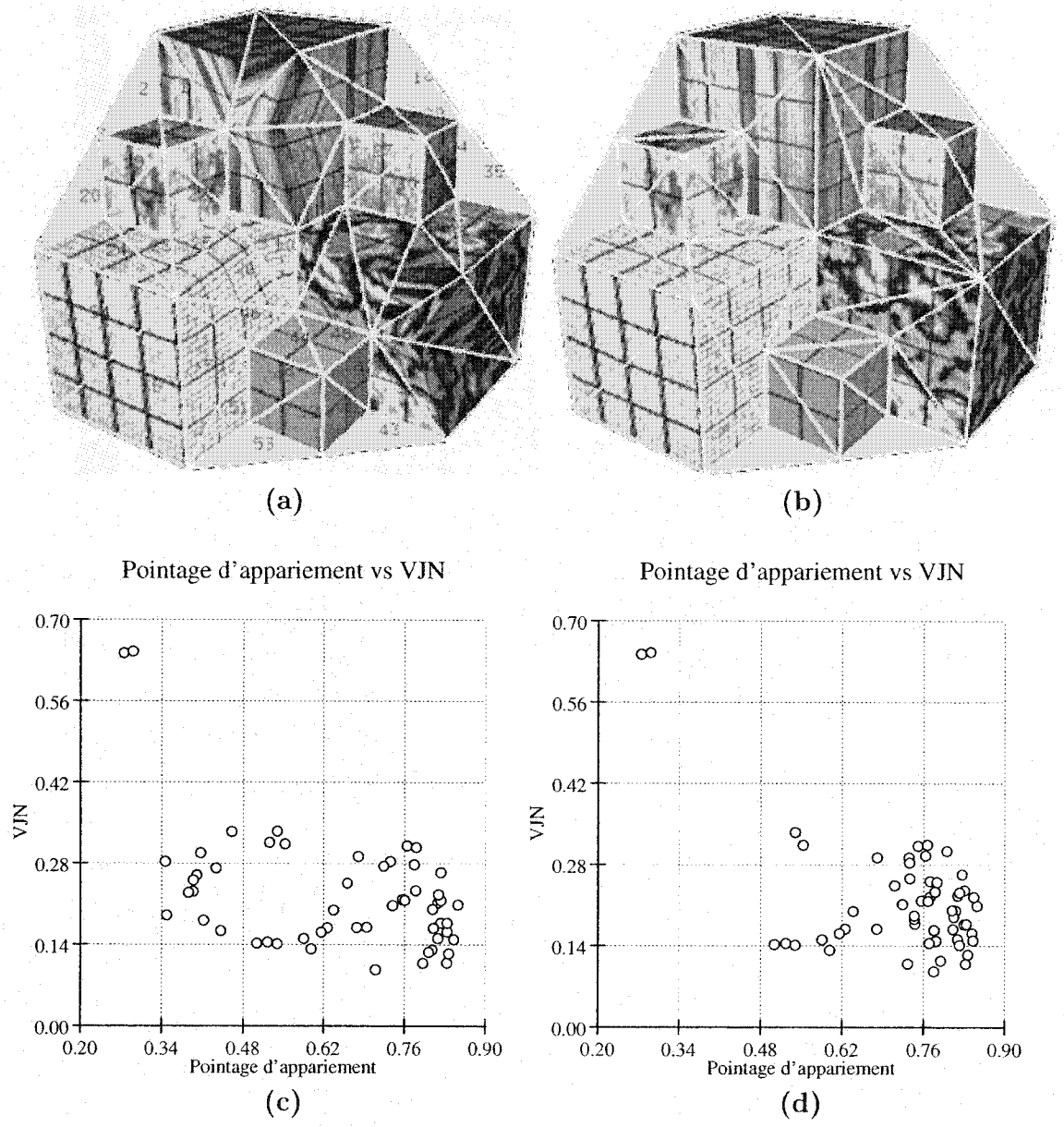


Figure 5.9 : Résultats de la re-triangulation pour la scène montrée dans les figures 5.31-a et 5.31-b. **(a)** Triangulation de Delaunay initiales des points d'appariement. **(b)** Triangulation physiquement valide générée par l'algorithme proposé. **(c)–(d)** Mesures de similarité et de fiabilité pour chaque triangle avant et après la re-triangulation, respectivement.

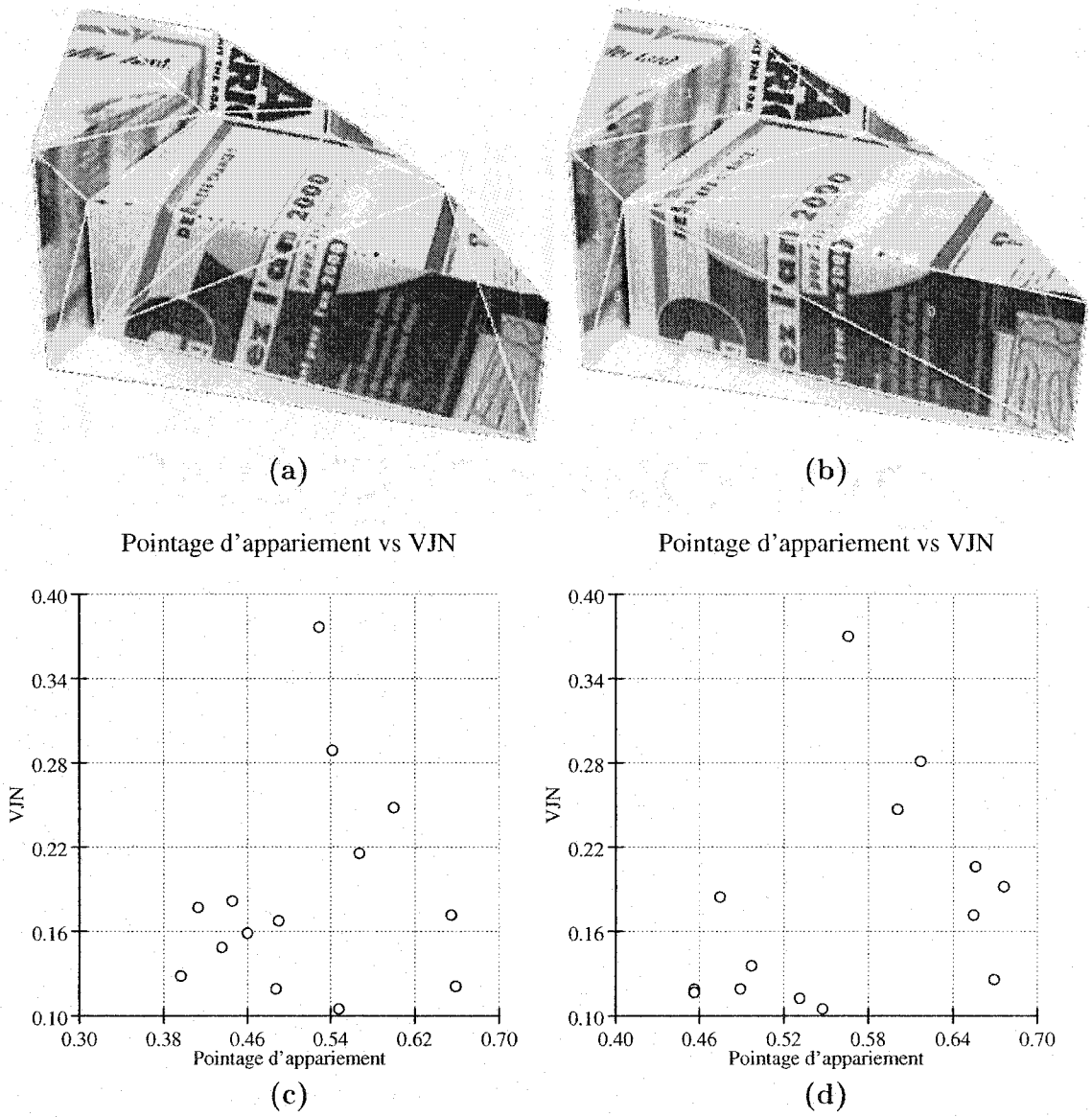


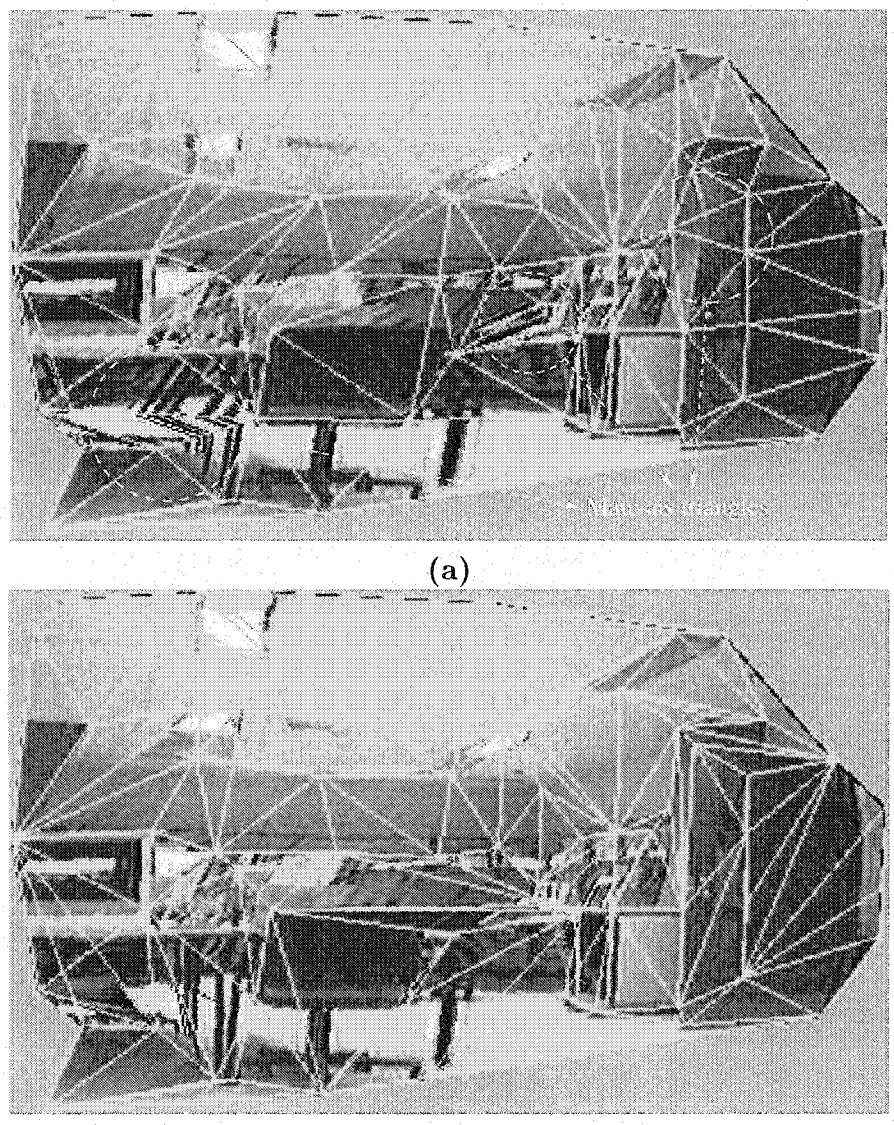
Figure 5.10 : Résultats de la re-triangulation pour la séquence des boîtes en rotation. **(a)** Triangulation de Delaunay initiales des points d'appariement. **(b)** Triangulation physiquement valide générée par l'algorithme proposé. **(c)–(d)** Mesures de similarité et de fiabilité pour chaque triangle avant et après la re-triangulation, respectivement.

résultats, dans le même format que la section précédente, sont montrés dans la figure 5.12. La figure 5.12-(a) montre l'image rectifiée résultant de la triangulation de Delaunay. L'utilisation de cette triangulation pour la synthèse de vue produit des résultats insatisfaisants. Certains des triangles qui dérangent le plus la compréhension et la visualisation de la scène pour la synthèse de vue sont encerclés. Dans la figure 5.12-(b), ces triangles ont été corrigés. Les figures 5.13-(a) et (b) sont les graphiques des pointages de similarité et de fiabilité pour les triangles. L'effet de la re-triangulation n'est pas clairement visible, mais on peut quand même remarquer que la moyenne des pointages de similarité a augmentée légèrement de 0.735928 à 0.753780 ce qui démontre une amélioration de la triangulation. De plus, la scène ne contient pas beaucoup d'information visuel pour faire l'évalutation de similarité. Ceci est démontré dans les figures 5.13-a et 5.13-b, où la grande majorité des triangles ont un pointage VJN en dessous de 0.1, ce qui est globalement très bas. Les algorithmes de mise en correspondance en général, ne fonctionnent pas bien sur ce genre d'image. Malgrés tout, l'algorithme de re-triangulation est parvenu à produire une triangulation améliorée par rapport à une simple triangulation de Delaunay.

5.3 Raffinement de la triangulation

L'approche de triangulation décrite dans ce travail vise à améliorer les images synthétisées en couplant la triangulation avec une interprétation géométrique de la scène. Cette interprétation géométrique est basée sur la maximisation de la mesure de corrélation basée sur la texture. Afin d'évaluer la pertinence de l'utilisation de cette mesure de corrélation en tant qu'estimation de l'erreur d'interprétation géométrique, une expérience a été conduite sur une scène synthétique (montrée dans les figures

Figure 5.11 : Images de la scène du laboratoire. La caméra a subit une rotation autour de l'axe vertical entre les deux images.



(a)

(b)

Figure 5.12 : Résultats de la re-triangulation pour la scène du laboratoire. **(a)** Triangulation de Delaunay initiales des points d'appariement. **(b)** Triangulation physiquement valide générée par l'algorithme proposé.

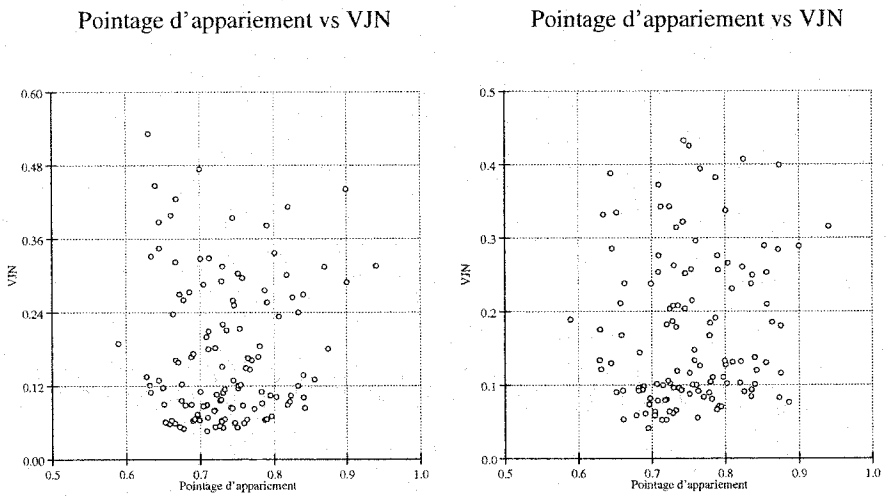


Figure 5.13 : Résultats de la re-triangulation pour la scène du laboratoire. **(a)–(b)** Mesures de similarité et de fiabilité pour chaque triangle avant et après la re-triangulation, respectivement.

5.31-a et 5.31-b) dans laquelle l'approximation géométrique a été évaluée en utilisant l'information contenue dans la carte de profondeur (z-buffer) qui a été produite lors du rendu de la scène.

Étant donné que les triangles sont des surfaces planes, la valeur de la profondeur des sommets d'un triangle est utilisée pour interpoler la profondeur de chaque point à l'intérieur du triangle. Ces valeurs sont ensuite comparées aux valeurs correspondantes dans la carte de profondeur et l'erreur est calculée tel que montré dans la figure 5.14.

L'approximation de la scène produite par le modèle composé de triangles ainsi que l'évaluation de l'erreur de profondeur associée pour plusieurs itérations de raffinement sont présentées dans la figure 5.24. La colonne de gauche présente la transformation des triangles de la première image sur les triangles de la seconde image. La colonne du centre montre la triangulation superposée à l'image transformée. Finalement, la

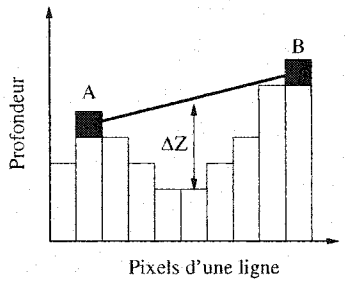


Figure 5.14 : Mesure de l'erreur géométrique pour un triangle donné en utilisant la carte de profondeur. La profondeur à l'intérieur du triangle est interpolée en se basant sur la profondeur des sommets. La différence ΔZ entre les valeurs interpolées et les valeurs connues à chaque pixel est mesurée.

colonne de droite affiche l'erreur sur la profondeur de l'approximation triangulaire. Les rangées (du haut vers le bas) sont les itérations 0, 44, 109 et 207. On peut observer que l'erreur sur la profondeur est considérablement réduite avec la progression des itérations de l'algorithme de raffinement proposé. On peut noter que dans ce cas, la triangulation initiale de Delaunay produit de grandes erreurs sur la profondeur à cause de l'absence de points d'appariement sur les coins des cubes. Ces coins ont été ensuite ajoutés automatiquement par l'algorithme, à l'itération 44 (seconde rangée dans la figure). Un exemple similaire d'une scène avec des surfaces plus lisses est présenté dans la figure 5.20.

5.4 Séquences de raffinement

Cette section contient une série de tests complets sur quatre scènes. Chaque scène est utilisée pour illustrer le processus de raffinement de la triangulation.

La première scène, montrée dans la figure 5.15, est une scène virtuelle composée d'une surface lisse avec des ondulations aléatoires. Les variations de la surface ont des amplitudes et tailles différentes afin d'observer le comportement de l'algorithme

de raffinement. Puisque cette scène est virtuelle, la carte de profondeur est également disponible. Le mouvement de la caméra entre les deux images est un léger mouvement vertical vers le sommet de l'objet. La figure 5.16 montre quatre états de la triangulation durant le processus de raffinement. Chaque rangée correspond à un état de la triangulation. La première colonne contient l'image originale. La seconde colonne contient l'image transformée (rectifiée). La troisième colonne contient la triangulation superposée à l'image rectifiée. Et finalement, la quatrième colonne contient l'erreur en profondeur entre la triangulation et la carte de profondeur. La figure 5.17 illustre les résultats comparatifs de l'utilisation de la triangulation pour produire un modèle 3D de la scène afin de créer de nouvelles images. Premièrement, la méthode pour produire les modèles 3D n'est pas exact, il s'agit seulement d'une approximation proportionnelle de la scène en utilisant la disparité stéréo (la méthode est décrite dans la section 2.3.2.1). Les figures 5.17-(a) et (d) montrent l'objet original à partir des points de vue utilisés pour la synthèse d'image. La colonne du centre de la figure (b-e) montre les vues du modèle reconstruit en utilisant la triangulation physiquement valide alors que la colonne de droite montre le modèle reconstruit à partir de la triangulation de Delaunay. Il faut noter que les reconstructions contiennent le même nombre de triangles, donc, la différence entre les deux reconstructions n'est pas très prononcée. Finalement, la figure 5.18 contient les tracés des pointages globaux pondéré et moyen respectivement. On peut observer le même comportement au niveau du pointage basé sur l'image et le pointage basé sur la différence en Z pour les deux types de mesure (pondéré et moyen).

La figure 5.19 montre une autre scène virtuelle contenant une surface lisse avec une large dénivellation. Cette dénivellation est utilisée pour démontrer le comportement de l'algorithme de raffinement autour d'une grande erreur géométrique dans la

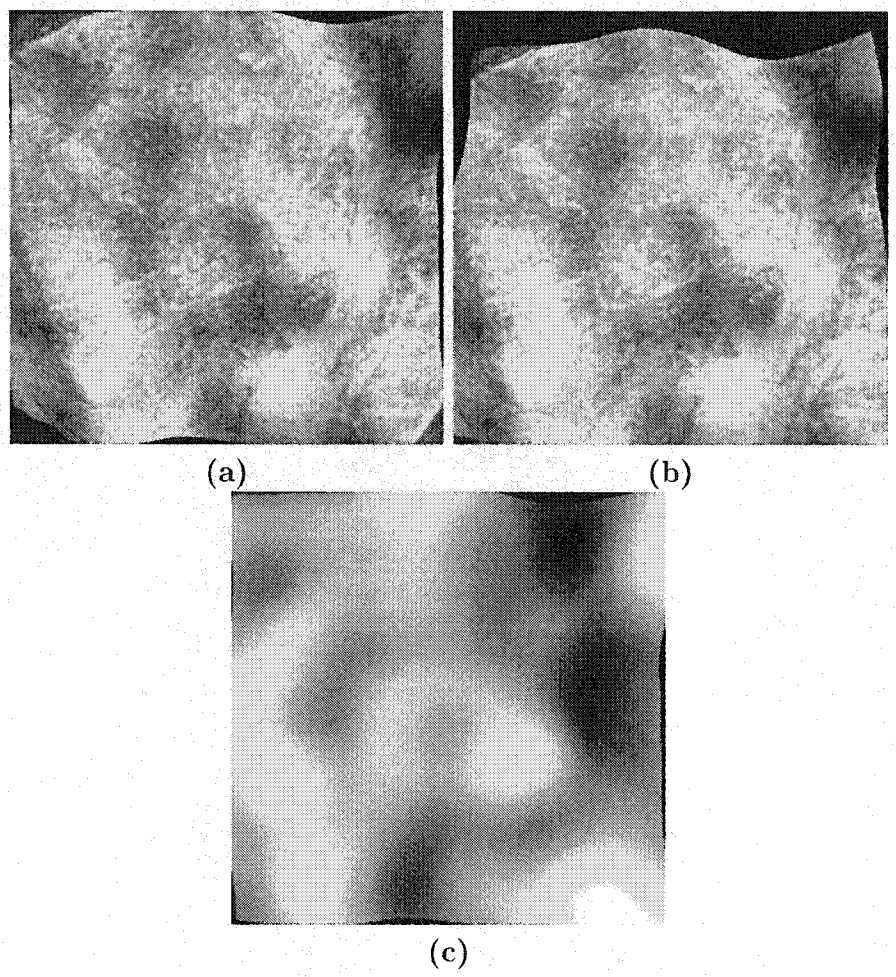


Figure 5.15 : Scène de test virtuel Surf0 - (a) Vue de la camera 1 (b) Vue de la camera 2 (c) Carte de profondeur

triangulation. Le processus de raffinement est montré dans la figure 5.20. La réduction de l'erreur en Z est visible tout au long de l'évolution de la triangulation. Les résultats de synthèse de vue sont montrés dans la figure 5.21. Dans la colonne de droite, l'erreur dans la triangulation de la dénivellation est clairement visible. Cette erreur est grandement réduite dans la colonne du centre en utilisant la triangulation P.V. (physiquement correcte). Une fois de plus, les deux triangulations ont le même nombre de triangles. Le modèle 3D créé à partir de la triangulation initiale, montré

dans la première rangée de la figure 5.20, n'a pas suffisamment de triangles pour créer un objet 3D reconnaissable. Le tracé des pointages est affiché dans la figure 5.22. Pour toutes les scènes, sauf celle de la figure 5.23, les points d'appariement ont été obtenu automatiquement avec la méthode décrite dans (Zhang et al., 1994).

La figure 5.23 montre la scène virtuelle composée d'un empilement de boîtes. L'évolution du processus de raffinement est illustrée dans la figure 5.24. On peut voir que l'erreur autour des coins manquants est réduite graduellement. Les modèles 3D créés à partir des triangulations physiquement valides et de Delaunay sont montrées dans la figure 5.25. Finalement, les tracés des résultats de pointages globaux pondéré et moyen sont affichés dans la figure 5.26.

Le dernier exemple est une scène réelle prise d'une séquence d'images. La scène est montrée dans la figure 5.27. La carte de profondeur pour cette image est créée à partir d'une carte de disparité obtenue par l'algorithme de croissance de région décrit dans la section 2.2.2.2. On peut voir deux cartes de profondeur obtenue en utilisant des méthodes de filtrage légèrement différentes sur la carte de disparité. La séquence de raffinement, montrée dans la figure 5.28 est faite avec la carte de profondeur afin d'illustrer l'erreur en Z. Puisque la surface de la roche n'est pas lisse, il est difficile d'évaluer avec précision l'évolution de l'erreur. Par contre, on peut remarquer que l'erreur diminue globalement. La figure 5.29 illustre deux modèles de la roche à partir de point de vue différent. Finalement, les tracés des pointages globaux sont montrés dans la figure 5.30. Étant donné que la carte de profondeur contient une estimation de la vraie forme de la roche, les tracés d'erreur de profondeur devraient être considérés comme largement imprécis. Néanmoins, cette estimation donne tout de même une idée de l'évolution de l'erreur et semble suivre le pointage de similarité.

Le comportement de la mesure d'erreur sur la profondeur comparée à la mesure

de planarité pour les itérations de la figure 5.24 est présenté dans la figure 5.26. Les figures 5.26-a et 5.26-c présentent les pointages globaux pondérés S_w et moyen S_m tel que définit dans (4.7). Les figures 5.26-b et 5.26-d montrent les pointages pondéré et moyen de l'erreur de profondeur définie de la même façon que les pointages globaux.

On peut observer que la mesure de planarité est couplée à l'évaluation de l'erreur de profondeur, telle que désiré. En conséquence, il est montré expérimentalement que la mesure de planarité décrite dans la section 4.3 mesure de façon fiable la véritable erreur géométrique de la triangulation.

5.5 Résultats de synthèse de vues

Les effets d'une triangulation correcte peuvent être illustrés par la synthèse de vues en utilisant la méthode proposée. Des exemples de synthèse de vues sont montrés dans les figures 5.32 et 5.31. Les résultats affichés dans ces figures ont été synthétisés en se basant sur la reconstruction 3D des scènes montrées dans les figures 5.9 et 5.10. Les modèles 3d ont été créés en utilisant la disparité stéréo (voir section 2.3.2.1) de chaque paire de points appariés pour estimer la profondeur du point correspondant du modèle. La méthode proposée pour la triangulation physiquement valide a été utilisée pour produire les triangles reliant les points 3d du modèle.

Les figures 5.32-a et 5.31-c montrent une synthèse de vue qui est basée sur une reconstruction utilisant une triangulation de Delaunay. Comme on peut le voir, les triangles incorrects ne respectent pas la géométrie sous-jacente des objets de la scène.

Les figures 5.32-b et 5.31-d montrent des images synthétisées à partir de modèles reconstruits en utilisant la triangulation physiquement valide de la méthode proposée. Dans ce cas, tous les triangles respectent la géométrie de la scène.

La triangulation physiquement correcte permet l'usage de méthodes de reconstruction 3d, qui, ensuite, peuvent être utilisées avec les cartes graphiques actuelles pour accomplir de la synthèse de vue en temps réel.

L'algorithme de triangulation itératif proposé est une solution flexible au problème de synthèse de vue. Il est possible de générer de nouvelles vues de la scène basées sur une triangulation initiale plus grossière et ensuite, raffiner les images lorsque la triangulation est graduellement raffinée. Une démonstration de cette propriété est donnée dans la figure 5.33. Les vues originales sont montrées dans les figures 5.33-a et 5.33-b. Les triangulations initiales et raffinées sont affichées dans les figures 5.33-c et 5.33-d. Les points d'appariements ont été obtenus automatiquement par la méthode décrite dans (Zhang et al., 1994). Les résultats d'une synthèse de vue basée sur les modèles 3D reconstruits à partir des triangulations des figures 5.33-c et 5.33-d sont présentés dans les figures 5.33-e et 5.33-f, respectivement. Comme on peut le voir, l'introduction de la triangulation raffinée réduit les distorsions dans les images synthétisées et augmente le niveau de détail. Le processus de raffinement peut être arrêté lorsque le niveau de détail est suffisant pour la tâche à accomplir. Les résultats d'une synthèse de vue basée sur l'interpolation linéaire de la même triangulation sont présentés dans les figures 5.33-g – 5.33-h, respectivement. L'interpolation de la triangulation initiale est embrouillée parce que les triangles correspondants ne contiennent pas la même texture. Par contre, la triangulation raffinée est claire et les détails de la roche sont visibles.

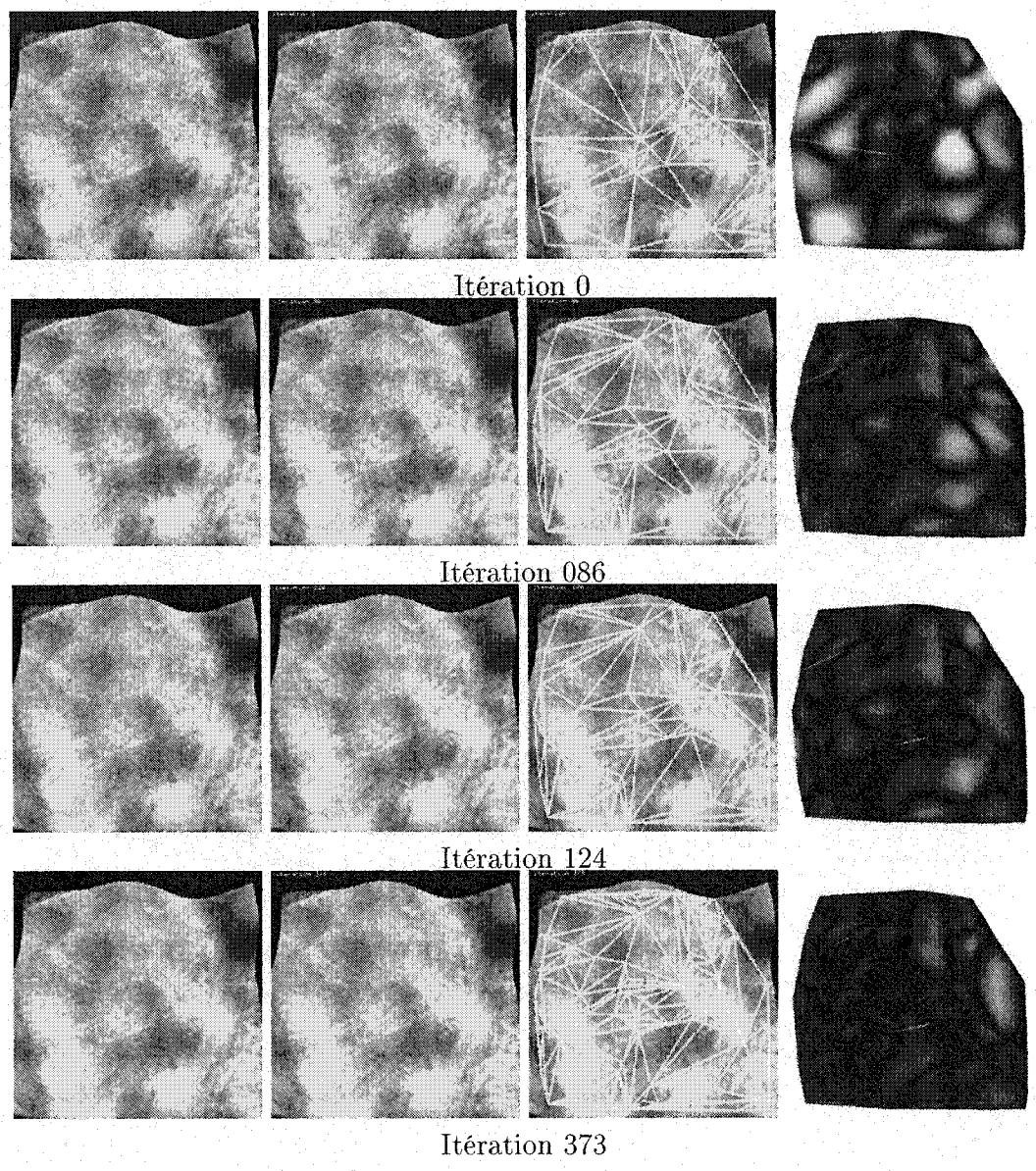


Figure 5.16 : Évolution de la scène "Surf0".

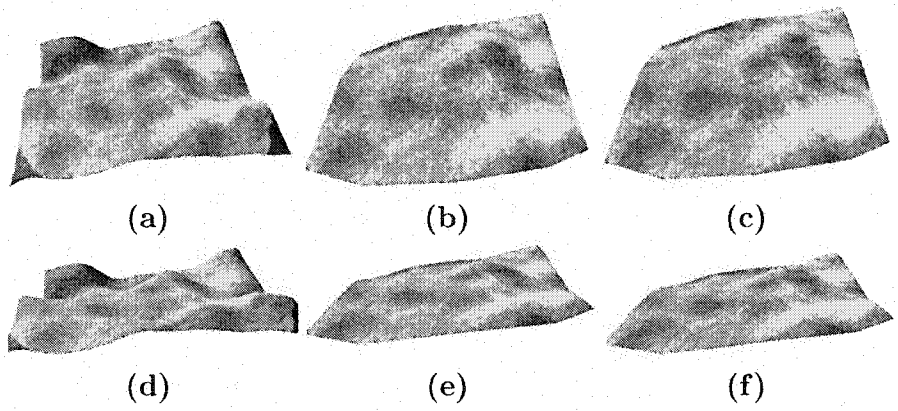


Figure 5.17 : Synthèse de vue sur la triangulation raffinée de Surf0 (88 points, 156 triangles) - **(a)** Vue 0 de la vraie surface. **(b)** Vue 0 de la triangulation PV **(c)** Vue 0 de la triangulation de Delaunay **(d)** Vue 1 de la vraie surface **(e)** Vue 1 de la triangulation PV **(f)** Vue 1 de la triangulation de Delaunay

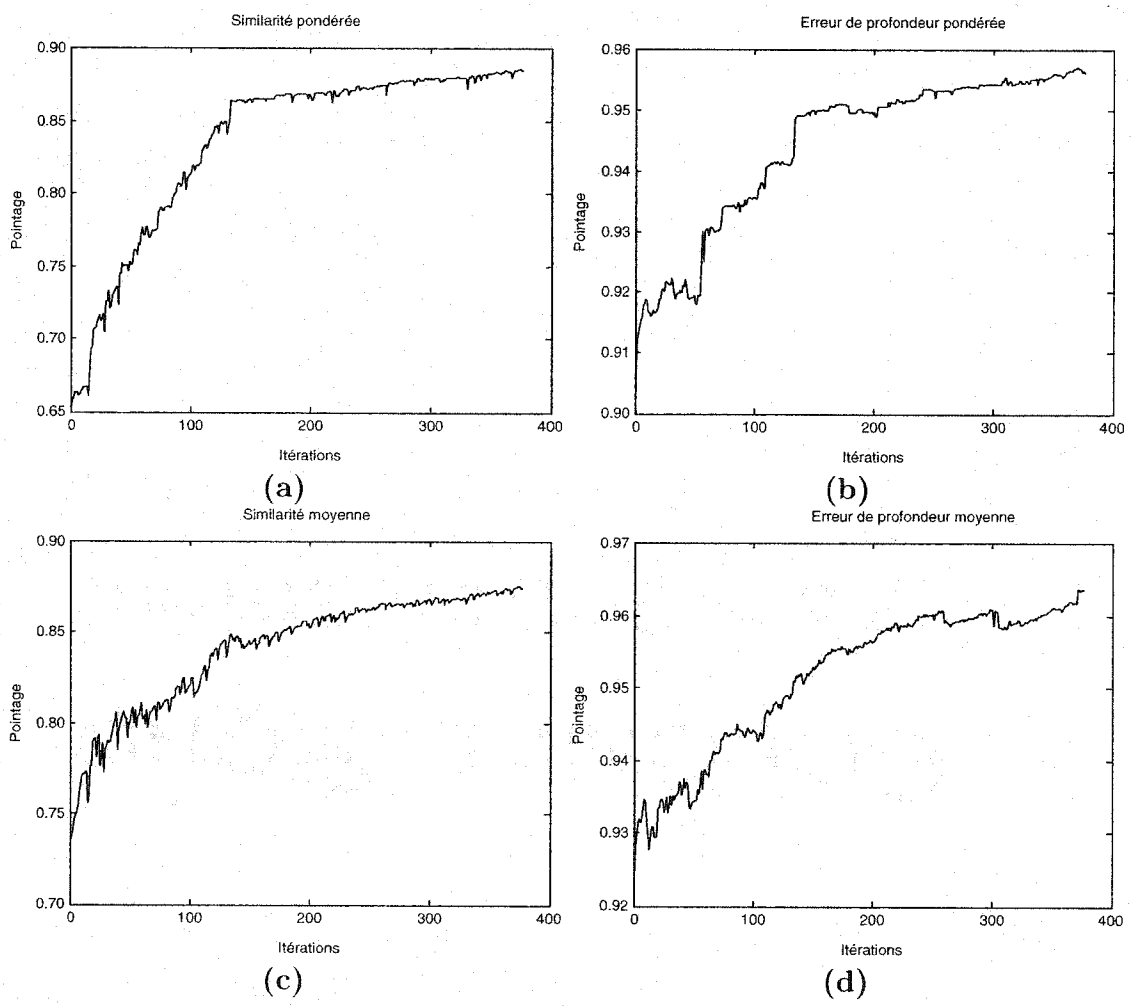


Figure 5.18 : Tracés des résultats pour la scène Surf0 **(a)** Pointage de similarité pondéré **(b)** Pointage d'erreur en Z pondéré **(c)** Pointage de similarité moyen **(d)** Pointage d'erreur en Z moyen

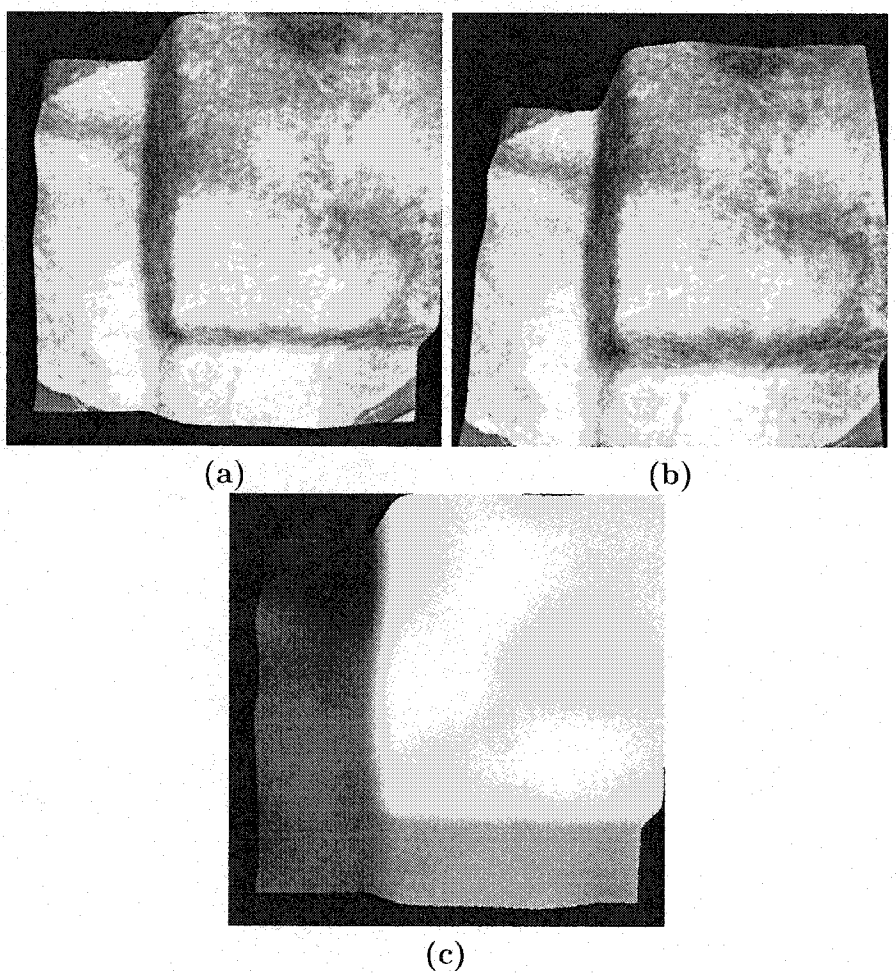


Figure 5.19 : Scène virtuelle Edge0 - (a) Vue de la camera 1 (b) Vue de la camera 2 (c) Carte de profondeur

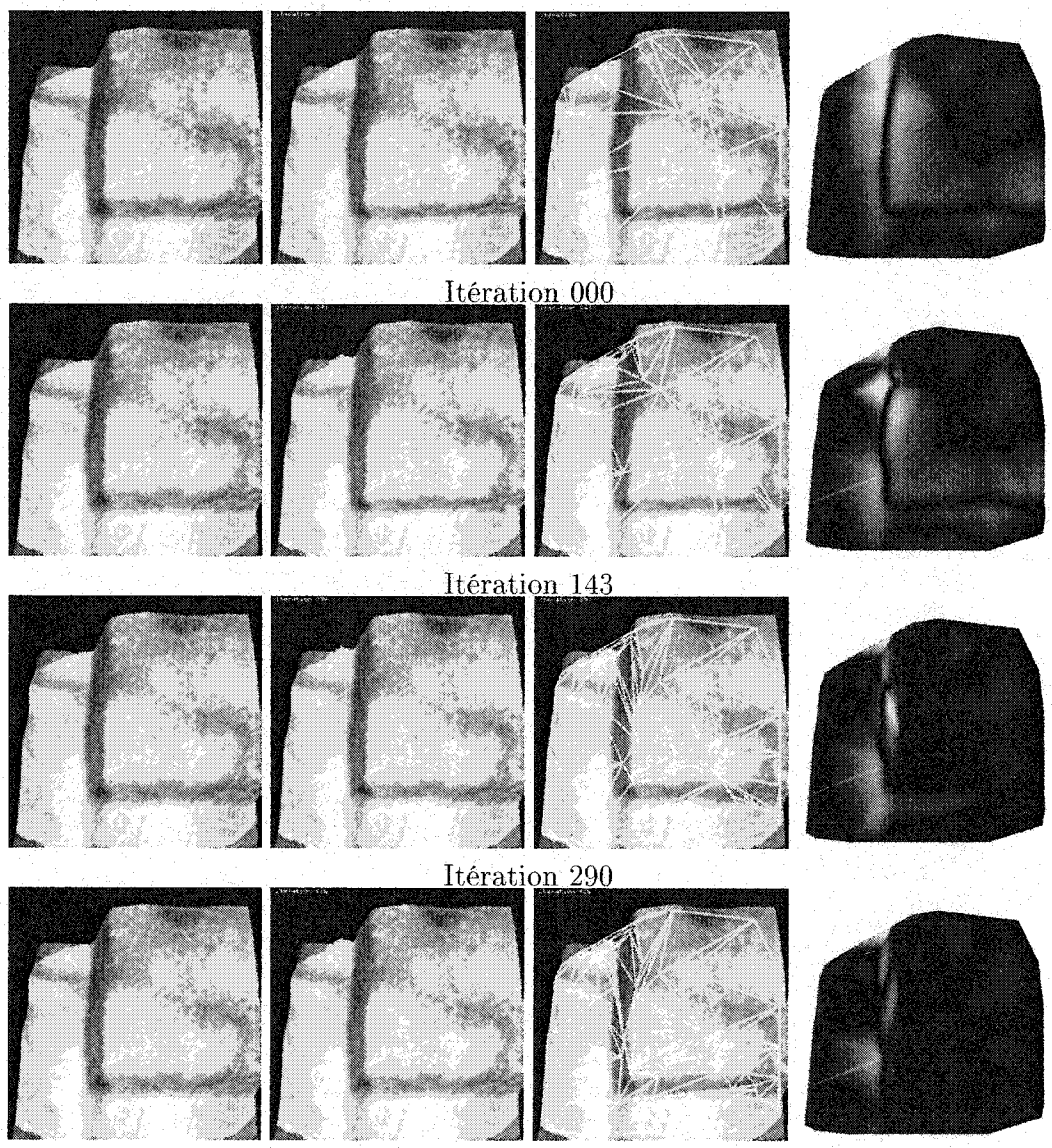


Figure 5.20 : Évolution de la triangulation.

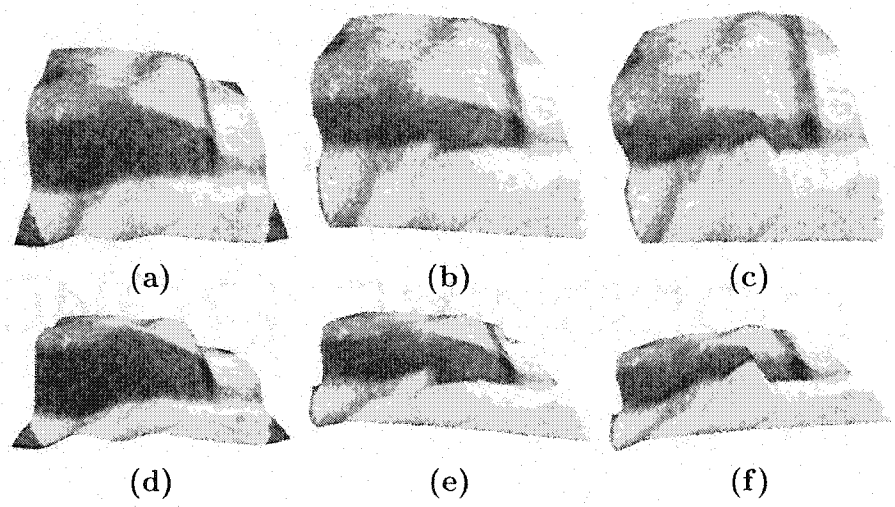


Figure 5.21 : Synthèse de vue sur la triangulation raffinée de Edge0 (88 points, 156 triangles) - **(a)** Vue 0 de la vraie surface **(b)** Vue 0 de la triangulation PV **(c)** Vue 0 de la triangulation de Delaunay **(d)** Vue 1 de la vraie surface **(e)** Vue 1 de la triangulation PV **(f)** Vue 1 de la triangulation de Delaunay

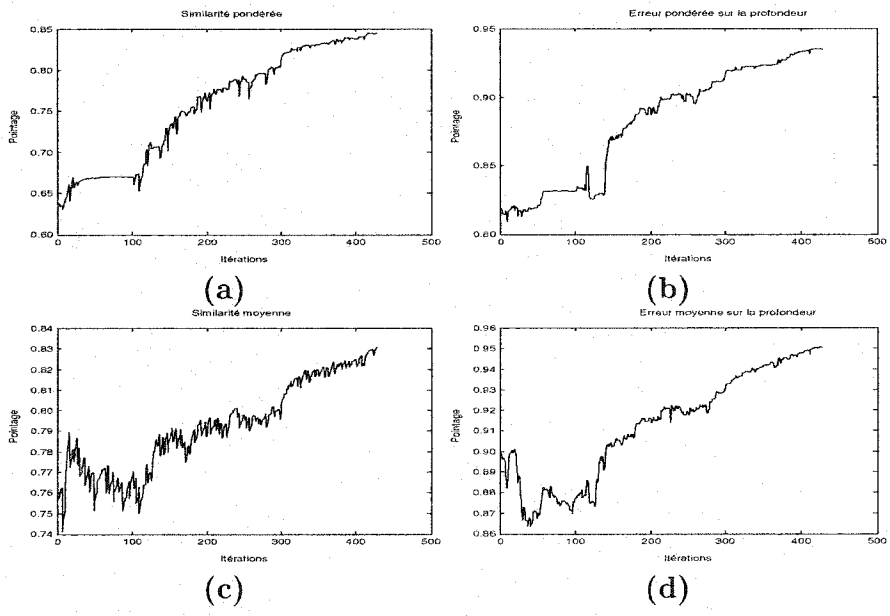


Figure 5.22 : Tracés des résultats pour la scène Edge0 **(a)** Pointage de similarité pondéré **(b)** Pointage d'erreur en Z pondéré **(c)** Pointage de similarité moyen **(d)** Pointage d'erreur en Z moyen

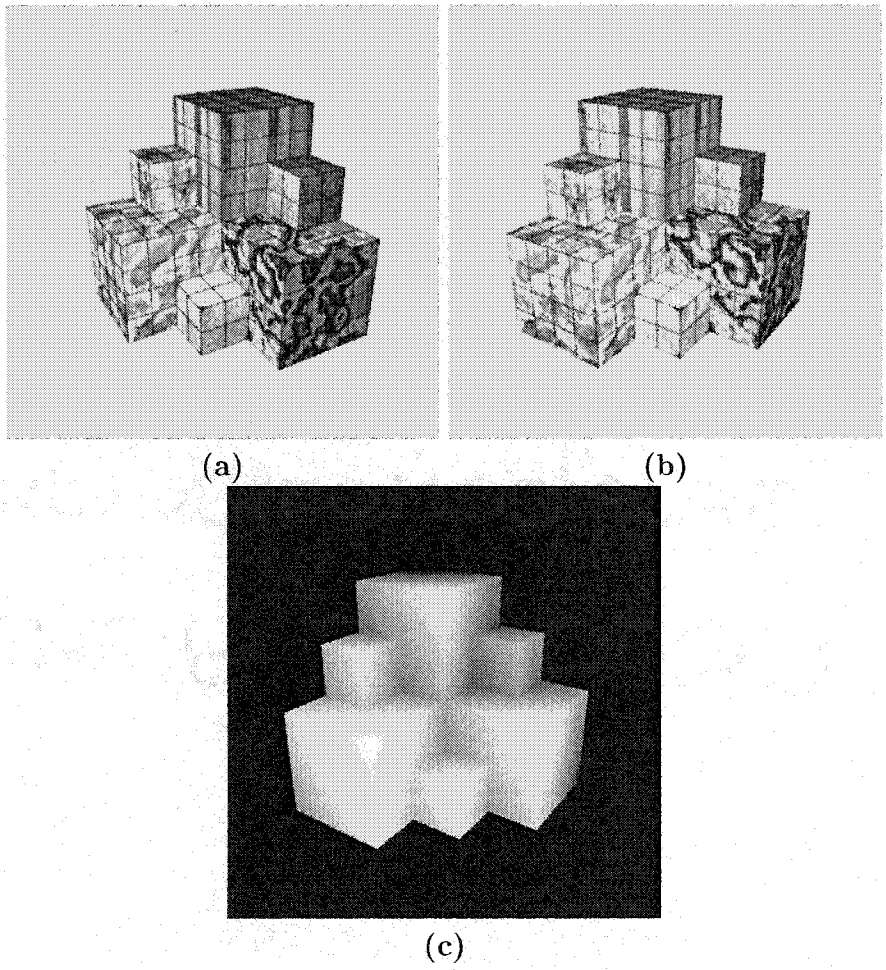
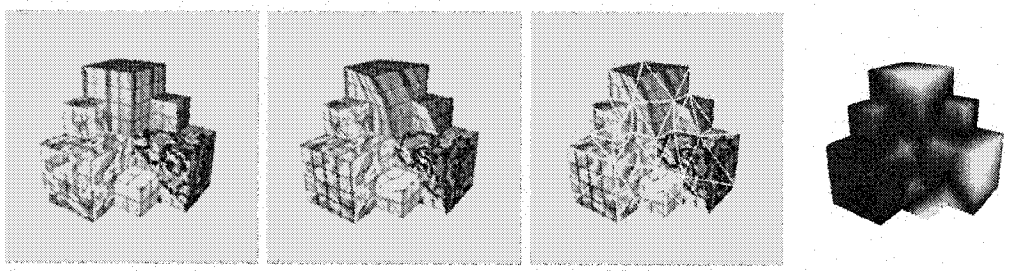
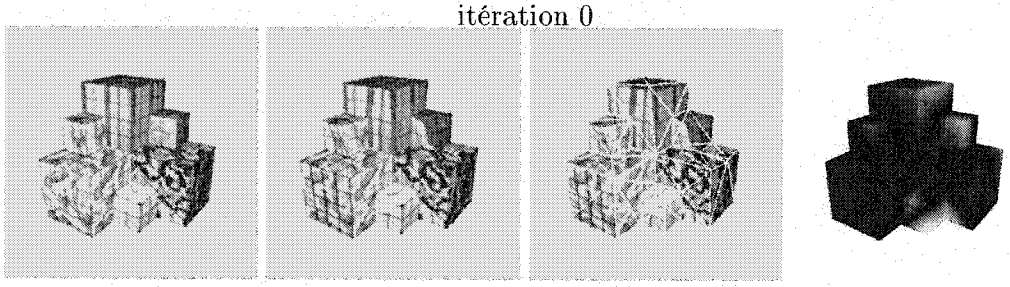


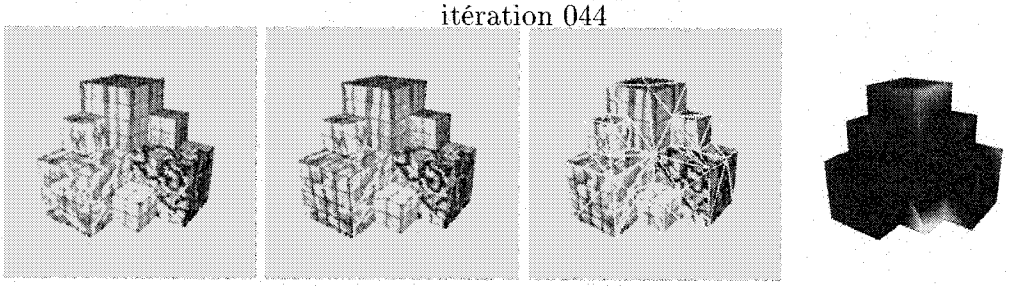
Figure 5.23 : Scène virtuelle "vboxgrp" - **(a)** camera 1 **(b)** camera 2 **(c)** carte de profondeur



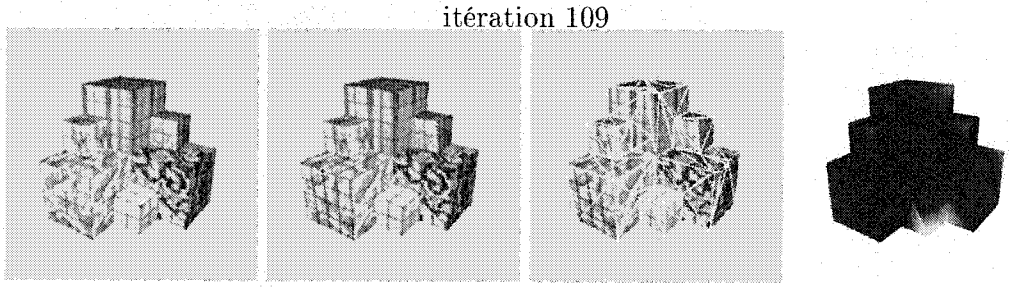
itération 0



itération 044

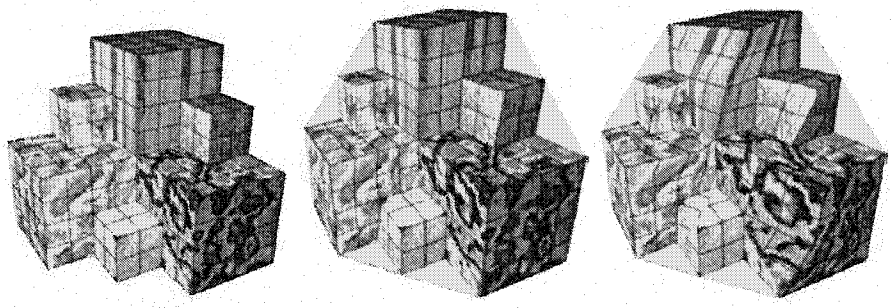


itération 109



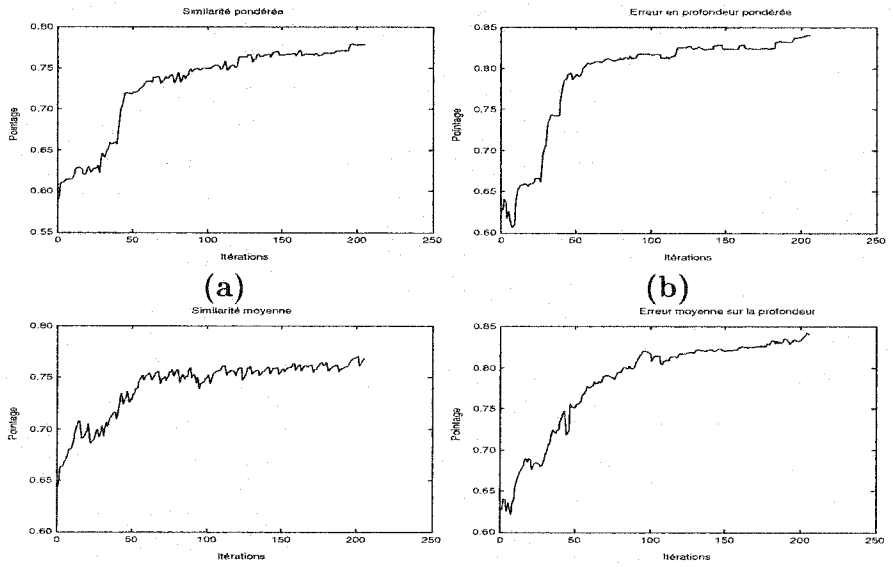
itération 207

Figure 5.24 : Évolution de la triangulation.



(a) (b) (c)

Figure 5.25 : Synthèse de vue sur la triangulation rafinée de la scène vboxgrp (55 points, 98 triangles) - **(a)** scène réelle **(b)** triangulation P.V. **(c)** triangulation de Delaunay



(a) (b) (c) (d)

Figure 5.26 : Tracés des résultats pour la scène *vboxgrp* **(a)** Pointage de similarité pondéré **(b)** Pointage d'erreur en Z pondéré **(c)** Pointage de similarité moyen **(d)** Pointage d'erreur en Z moyen

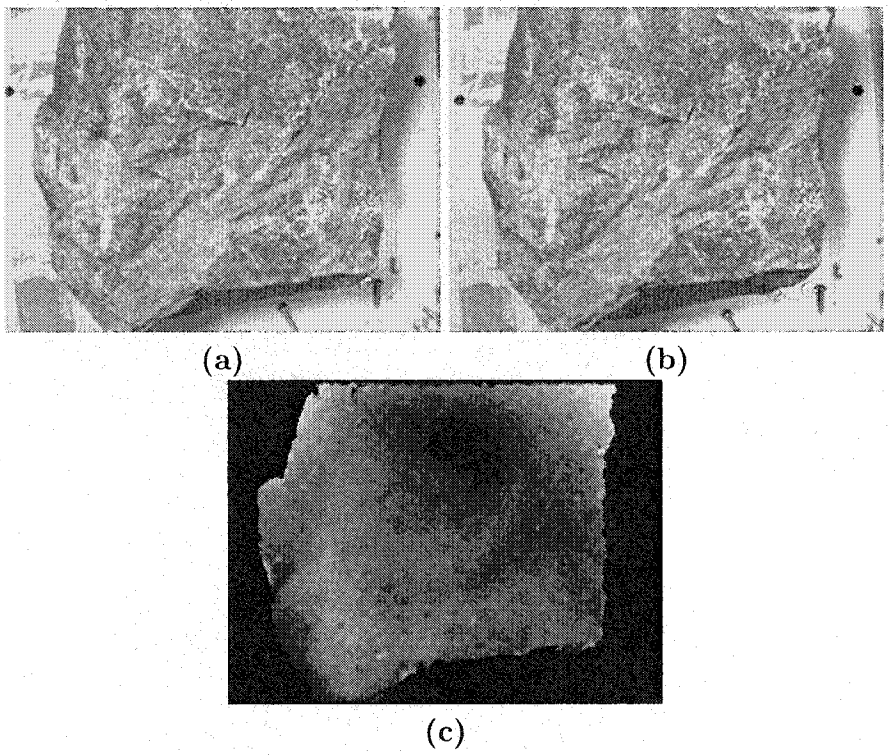


Figure 5.27 : Scène réelle de la roche - **(a)** Vue 0 **(b)** Vue 1 **(c)** Carte de profondeur (estimation avec la carte de disparité non rectifiée)

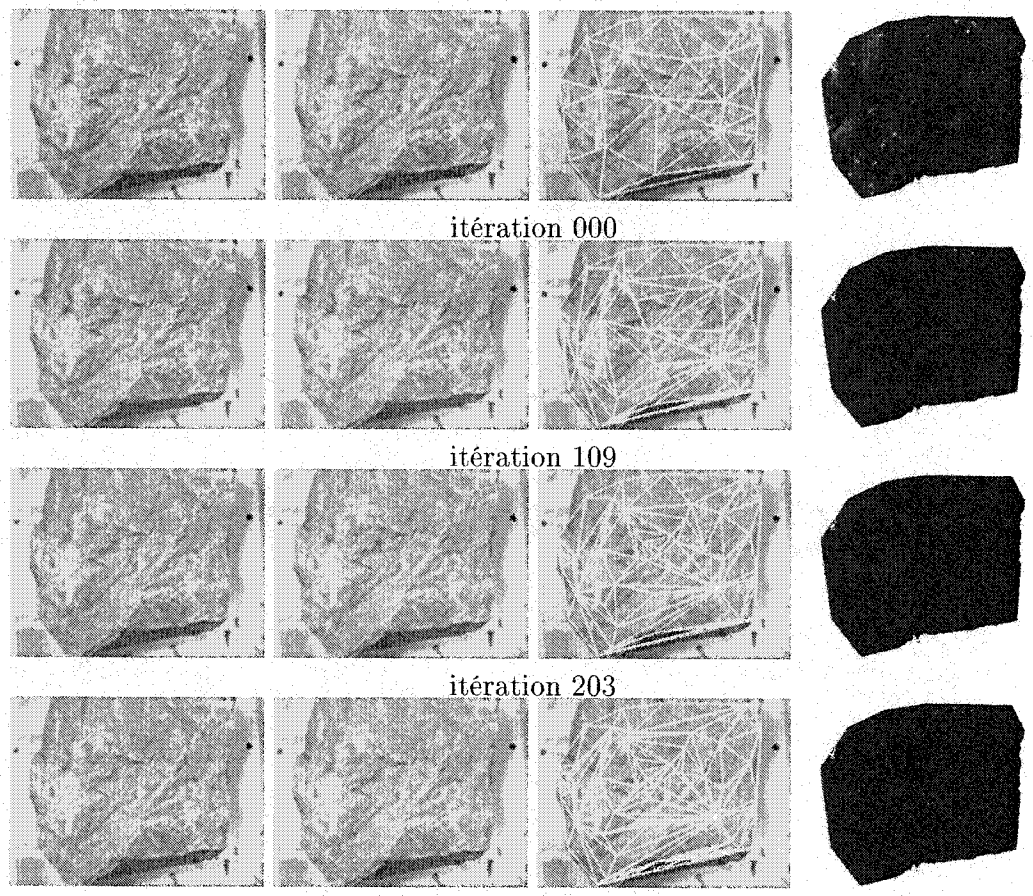


Figure 5.28 : Évolution de la triangulation.

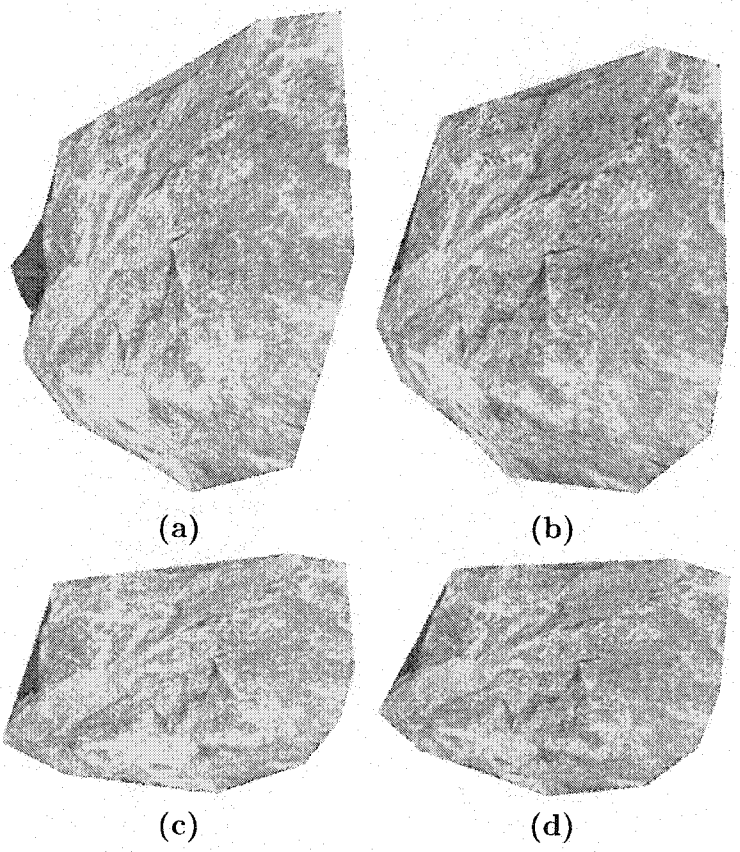


Figure 5.29 : Synthèse de vue de la triangulation raffinée pour la scène de la roche - **(a)** Vue 0 de la triangulation P.V. **(b)** Vue 0 de la triangulation de Delaunay **(c)** Vue 1 de la triangulation P.V. **(d)** Vue 1 de la triangulation de Delaunay

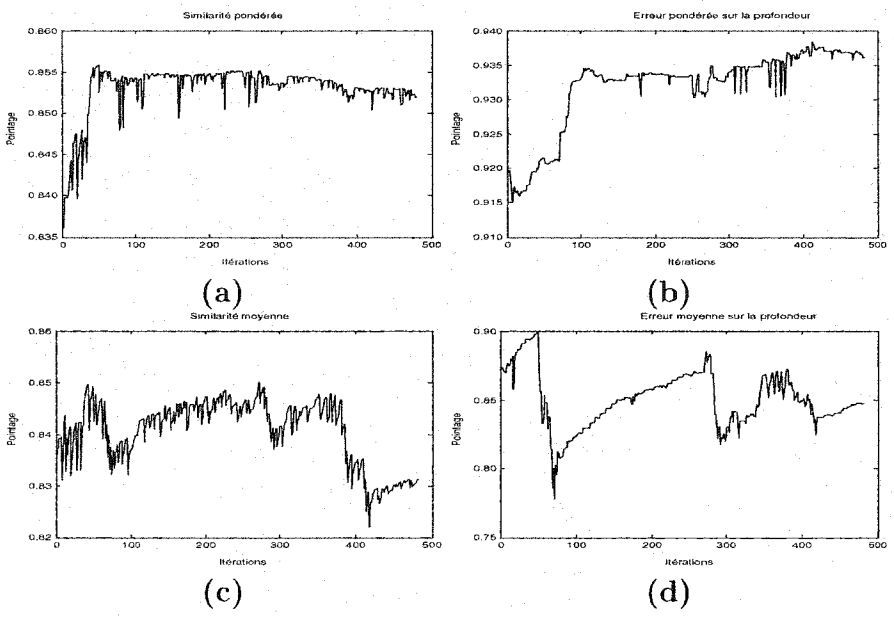


Figure 5.30 : Tracés des résultats pour la scène de la roche - **(a)** Pointage de similarité pondéré **(b)** Pointage d'erreur en Z pondéré **(c)** Pointage de similarité moyen **(d)** Pointage d'erreur en Z moyen

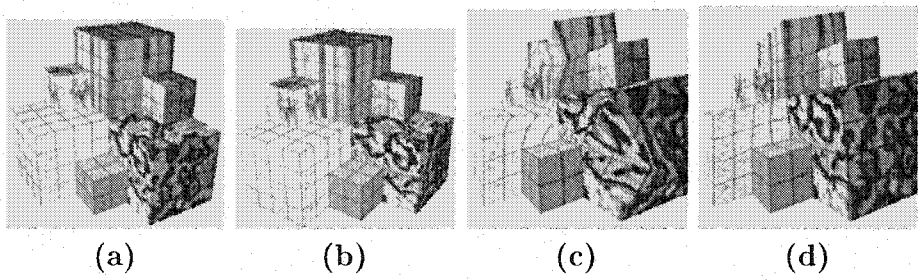
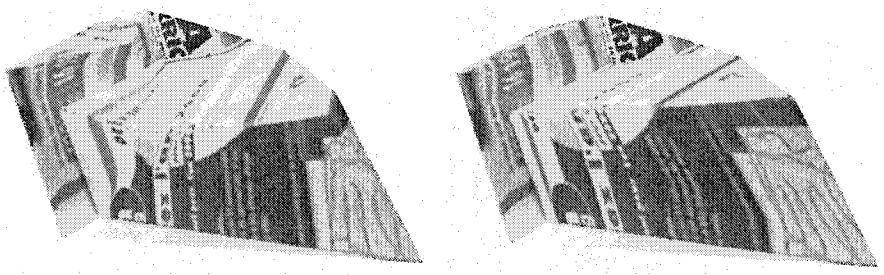


Figure 5.31 : **(a)**–**(b)** Image d'une scène virtuelle composée de boîtes. **(c)** Modèle 3d résultant de la triangulation de Delaunay. **(d)** Modèle 3d résultant de la triangulation physiquement valide.



(a) **(b)**

Figure 5.32 : **(a)** Modèle 3d résultant de la triangulation de Delaunay. **(b)** Modèle 3d résultant de la triangulation physiquement valide.

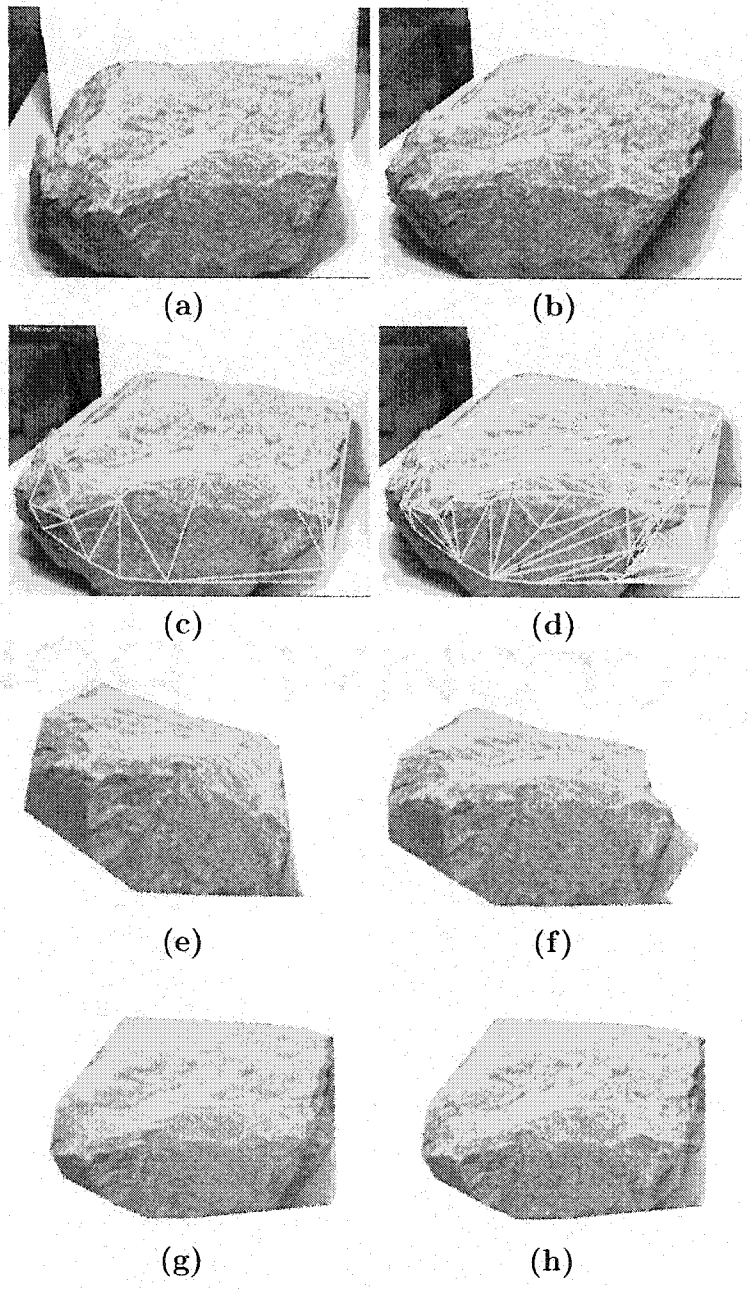


Figure 5.33 : Différentes approches pour la synthèse de vue de la roche. **(a)-(b)** Les images originales de la roche. **(c)-(d)** Triangulation initiale et raffinée, respectivement. **(e)-(f)** Reconstruction 3D des triangulations initiale et raffinée. **(g)-(h)** Interpolation d'images de la triangulation initiale et raffinée.

Chapitre 6

Résumé et conclusion

6.1 Résumé des contributions du mémoire

La synthèse de vues basée sur des images peut être effectuée de plusieurs façons, tel que présenté dans la section 3.1. Toutes ces méthodes utilisent des représentations (modèles) de la scène qui sont produites à partir des données disponibles. Le présent travail vise à trouver une façon d'extraire un modèle de la scène à partir d'un ensemble de points de correspondance épars entre deux images. Plusieurs méthodes satisfaisantes d'appariement épars existent déjà (voir (Zhang et al., 1994)). Le problème est donc de trouver une façon d'utiliser les points d'appariement et les images pour créer le modèle de la scène.

L'approche utilisée consiste à représenter la scène comme un ensemble de triangles interconnectés. Ces triangles sont une approximation de la forme de la scène réelle. Les surfaces courbes sont représentées par un plus grand nombre de triangles afin de minimiser l'erreur d'approximation. Pour obtenir un tel modèle de la scène, on utilise l'information disponible, c'est-à-dire, les points d'appariements ainsi que l'information

contenue dans les images. Il existe différentes façons de faire la triangulation de points. Une de ces façons est de faire une triangulation de Delaunay. On peut également y ajouter des contraintes supplémentaires en forçant certaines arêtes afin de conserver un contour déterminé dans la triangulation. Ces méthodes utilisent donc un critère uniquement géométrique pour déterminer la triangulation. Dans le cas de la synthèse de vue, c'est surtout la qualité de l'image résultante qui importe. Cela indique donc que c'est l'information contenue dans les images qui devrait déterminer la triangulation des points d'appariement. Pour ce faire, on établit un critère, ou une mesure, de qualité d'un triangle basé sur les images. Ce critère consiste à comparer le contenu des triangles correspondants dans les différentes images sources. Cette comparaison doit tenir compte de la projection perspective des surfaces selon leur point de vue. Ainsi, les triangles qui sont les projections d'une même surface plane (ou relativement plane selon la tolérance sur l'erreur de corrélation) auront une grande corrélation de l'image qu'ils contiennent. En se basant sur ce principe, on peut établir une mesure de la validité de chaque triangle. Ensuite, on utilise un opérateur de retriangulation, c'est-à-dire, le basculement d'arête, pour traverser l'espace des triangulations possibles. Il s'agit de maximiser la mesure de validité des triangles en utilisant l'opérateur de basculement d'arêtes pour passer d'une configuration à l'autre. On utilise une triangulation de Delaunay dans l'une des images sources comme point de départ de la recherche. Le choix de la triangulation de départ est relativement important. En effet, plus on prend un point de départ qui est "loin" de la triangulation recherchée et plus la recherche sera longue et le risque d'être coincé dans un maximum local sera élevé.

Même si l'on peut atteindre la triangulation parfaite avec les points d'appariements initiaux, il est toujours possible d'améliorer la qualité du modèle en ajoutant de nou-

veaux points d'appariements. L'ajout de ces points est dirigé par deux critères principaux. Premièrement, on veut une amélioration maximale du modèle. Ceci implique que les points seront ajoutés dans les régions où les triangles sont les plus erronés. Deuxièmement, on veut ajouter des points là où il y a de l'information disponible (des points d'intérêts) dans les images. Ces deux critères permettent de choisir judicieusement les triangles dans lesquels de nouveaux points seront ajoutés. Ceci permet d'accélérer la convergence de la méthode vers une triangulation de meilleure qualité. La méthode d'ajout de nouveaux points d'appariement à l'intérieur d'un triangle utilise le critère de validité des triangles. En effet, ce critère est le résultat de la comparaison du contenu des triangles. Plus le contenu des triangles est similaire, plus le résultat sera élevé. Ce critère est idéal pour déterminer la correspondance des sommets des triangles. Pour que le contenu de deux triangles soit similaire (corrélation élevée), il faut que chacun des sommets de ces triangles soit un point d'appariement valide. Ainsi, à l'aide du critère de validité des triangles, on peut en déduire la correspondance de leurs sommets. La méthode de raffinement de la triangulation utilise ce principe pour l'ajout des nouveaux points d'appariement.

Toujours afin d'améliorer le modèle de la scène, la méthode de raffinement de la triangulation doit être capable d'éliminer les points d'appariement erronés. Ces points peuvent provenir de l'ensemble initial ou bien avoir été ajoutés par la méthode de raffinement. Encore une fois, on utilise le critère de validité des triangles pour tenter de trouver les sommets invalides. En principe, lorsqu'un point d'appariement est incorrect, il va rendre les triangles auxquels il appartient incorrects également. On peut ainsi déduire une mesure de validité d'un sommet par la validité des triangles qui y sont connectés. Par contre, il est possible qu'un sommet soit un point d'appariement valide, mais que tous les triangles qui y sont connectés soient invalides.

C'est pourquoi un point de correspondance est éliminé seulement si, ce faisant, on permet à la triangulation de s'approcher de la solution optimale. La méthode consiste à enlever un point "suspect", refaire la triangulation jusqu'à atteindre un maximum local de la qualité du modèle de la scène. On le compare ensuite avec le maximum local précédent l'élimination du point "suspect". Le nouvel état de la triangulation est conservé seulement si son maximum local est supérieur à celui de l'état précédent.

La méthode de triangulation des points est donc une combinaison des étapes de re-triangulation par basculement d'arêtes, de raffinement par ajout de nouveaux points et élimination des points erronés. L'ajout de nouveaux points et l'élimination des points erronés permettent de se déplacer d'avantage dans l'espace des triangulations possibles et également d'éviter les maximums locaux dans une configuration donnée. En effet, lorsque l'algorithme de re-triangulation atteint un maximum local, l'ajout de nouveaux points et l'élimination des points erronés vont permettre d'évoluer au-delà de ce maximum local et de s'approcher d'une meilleure triangulation. Cette nouvelle approche pour la triangulation est également décrite dans (Perrier et al., 2000*a*; Perrier et al., 2000*b*).

L'utilisation d'un ensemble de triangles comme représentation intermédiaire dans un processus itératif a plusieurs avantages. La triangulation est directement utilisable pour la synthèse de vue. Aussitôt qu'une première triangulation est disponible, les données peuvent être utilisées pour produire de nouvelles images de la scène. Ensuite, lorsque les itérations progressent, les nouvelles vues de la scène deviennent plus précises et plus détaillées. Le processus de raffinement utilise également les données directement, ce qui veut dire qu'il n'y a pas de conversion de donnée entre chaque tâche (raffinement et génération de nouvelles vues).

6.2 Limites et contraintes

La triangulation d'un ensemble produit toujours une forme qui correspond à l'enveloppe convexe des points. Puisque l'algorithme de raffinement ajoute de nouveaux points seulement à l'intérieur des triangles déjà existants, la région couverte par la triangulation ne grandira jamais au-delà de l'enveloppe convexe des points d'appariement initiaux. Ceci suggère la création d'un nouvel algorithme qui utiliserait la mesure de similarité des triangles afin d'ajouter des points, et donc des triangles, à l'extérieur de la triangulation actuelle. Ceci pourrait être fait de façon similaire à l'algorithme de raffinement, en détectant des points d'intérêts (coins) dans la région d'arrière-plan à l'extérieur de l'enveloppe convexe. La région de recherche devrait cependant être limitée. Par exemple, la fenêtre de recherche pourrait être limitée à un quart de l'image autour du point considéré. Cette méthode deviendrait similaire à la méthode de croissance de région (décrit dans la section 2.2.2.2). L'ajout de nouveaux points à l'extérieur de l'enveloppe convexe pourrait ajouter plus d'un triangle à la fois, donc, le pointage de similarité globale devrait être utilisé dans le critère de rejet ou d'acceptation des nouveaux points.

Une autre limite à l'utilisation de la triangulation comme modèle de scène pour la synthèse de vue est l'incapacité d'utiliser la région hors de l'enveloppe convexe pour faire le rendu des images. Par exemple, si tous les points d'appariements sont sur un seul objet, alors, seulement cet objet peut être utilisé pour la synthèse de vue. L'arrière-plan, qui est le reste de l'image, ne sera pas utilisé puisqu'il n'est pas dans la région appariée couverte par la triangulation. De plus, les régions qui peuvent être utilisées comme source d'information doivent absolument être visibles dans les deux images sources. Donc, les parties de la scène qui sont visibles dans une seule image

ne peuvent pas être utilisées dans le modèle de la scène. Toutefois, ce fait est vrai pour n'importe quelle méthode de mise en correspondance et n'est pas une limite directement reliée à l'usage de la triangulation comme modèle intermédiaire. Par contre, il est possible, selon la méthode de synthèse de vue, d'utiliser des suppositions à propos de la région d'arrière-plan. Par exemple, on peut supposer que l'arrière-plan est une partie d'un plan ou d'un cylindre autour du point de vue à une distance prédéfinie. Alors, les régions des images sources contenant l'arrière-plan peuvent être utilisées pour faire une mosaïque qui servira d'arrière-plan pour la synthèse de vue. Ce genre de supposition sera toujours dépendante de la scène qui est observée et donc, une telle méthode ne pourrait pas s'appliquer en général. Des informations préliminaires sur la scène observée doivent être connues. Dans le cas de scènes complexes, comme des paysages, il n'est pas possible de trouver une supposition générale sur les zones d'arrière-plan puisque leur distance par rapport à l'observateur est difficile, sinon impossible, à prévoir.

Dans une scène complexe, où il y a beaucoup d'objets les uns en devant les autres, il y a souvent beaucoup d'occlusions. Ceci peut causer un problème lorsqu'un objet est visible dans une image et caché par un autre objet dans l'autre image. Dans ce cas, il n'est pas possible d'obtenir d'information de correspondance pour cet objet dans les zones occultées. Lorsque l'on utilise la triangulation, ces zones occultées seront couvertes par des triangles, mais ces triangles ne seront pas physiquement valides. Dans ce cas, il serait parfois préférable de simplement déconnecter ces triangles de leur voisin. L'évaluation de similarité peut être utilisée comme une pondération sur les connections des triangles pour produire un graphe de connectivité. Ensuite, en utilisant une méthode de groupement statistique, il serait possible d'identifier les liens les plus faibles du graphe et de les déconnecter. Ceci produirait des groupes de

triangles déconnectés correspondant aux objets occultés. Toutefois, cette méthode peut introduire des trous dans les images lorsque les triangles sont re-projetés dans de nouvelles vues. Le nombre et la taille de ces trous est hautement dépendant de la scène et de la position des points de vue originaux. Il est possible de réduire les trous en combinant les résultats de triangulation de plus de deux images, mais ce processus est complexe et est plus spécifique à la reconstruction 3D. Il n'y aurait tout de même pas de garantie que tous les trous seraient remplis correctement.

6.3 Nouvelles voies de recherche

La méthode proposée peut être adaptée à un grand nombre de techniques de synthèse de vues que ce soit de la simple interpolation d'image jusqu'à la reconstruction 3d complète. Elle peut être utilisée dans des systèmes de modélisation automatique ou semi-automatique ainsi que dans des variétés d'algorithmes graphiques telles que l'optimisation de modèle selon le point de vue. Dans des applications de réalité augmentée, où la position adéquate de modèle graphique dans une image est requise, le problème peut être simplifié considérablement en utilisant la méthode proposée pour créer un modèle de la scène basé sur le point de vue de l'observateur. Ce modèle pourrait ensuite être utilisé pour ajouter correctement des modèles virtuels à la scène.

L'algorithme de raffinement de la triangulation utilise des critères simples pour choisir les triangles qui seront fusionnés ou supprimés. D'autres critères pourraient être utilisés également tel que la forme de la scène et les occlusions potentielles. Toutefois, ceci introduirait des notions de reconstruction 3D implicite de la surface de la scène représentée par des triangles. Ceci implique donc que des informations supplémentaires, telles que la configuration de la géométrie épipolaire, soient connues.

De plus, dans l'algorithme de division des triangles, la correspondance est faite à partir des coins extraits de la texture des triangles correspondants. Ceci peut marcher correctement pour les images qui contiennent des coins, mais d'autres applications peuvent également utiliser des points d'intérêts spécifiques à leur besoin.

La méthode présentée dans ce travail gère la triangulation des points de correspondance entre deux images seulement. Cette méthode peut être utilisée avec plus de deux images. Dans un cas où plus de deux images sont disponibles, il serait possible d'utiliser une des images comme référence et rectifier toutes les autres images selon la référence. Ensuite, l'évaluation de planarité peut être accomplie sur les triangles correspondants de toutes les images. Ceci demanderait une généralisation de l'équation dans la section 4.3.1 afin de permettre la corrélation de plus de deux images à la fois. Ceci améliorerait grandement la fiabilité de la région de support pour l'évaluation de similarité et réduirait l'influence des différences de résolution entre les images. De plus, pour l'algorithme de division des triangles, dans le cas de trois images, le tenseur trilinéaire peut être utilisé pour réduire grandement la région de recherche des points d'appariements et augmenter la fiabilité des résultats de division.

Bibliographie

ADELSON, E. H. (1995), Layered representations for vision and video, dans *Proceedings IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA.

AGAM, G., MICHAUD, G., PERRIER, J. S., HOULE, J. L. et COHEN, P. (1999*a*), Image based view synthesis: a survey, *Submitted to Optical Engineering* .

AGAM, G., MICHAUD, G., PERRIER, J. S., HOULE, J. L. et COHEN, P. (1999*b*), A survey of image based view synthesis approaches for interactive 3D sensing, rapport technique GRPR-RT-9901, Perception and Robotics Laboratory, Ecole Polytechnique, Montreal, Canada.

AVIDAN, S., EVGENIOU, T., SHASHUA, A. et POGGIO, T. (1997), Image-based view synthesis by combining trilinear tensors and learning techniques, dans *VRST'97. ACM Symposium on Virtual Reality Software and Technology 1997*, Lausanne, Switzerland, pp. 103–10.

AVIDAN, S. et SHASHUA, A. (1997*a*), Novel view synthesis in tensor space, dans *Proceedings. 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 1034–1040.

AVIDAN, S. et SHASHUA, A. (1997*b*), Tensor embedding of the fundamental matrix, dans *Post-ECCV SMILE Workshop*, Vol. 1506.

BAKER, H. H. et BOLLES, R. C. (1989), Generalizing epipolar plane image analysis on the spatiotemporal surface, *International Journal of Computer Vision* **3**(1), 33–49.

BARRETT, E., PAYTON, P. M. et MARRA, P. (1997), Synthesizing perspective views of 3D scenes from multiple reference images using a synergy of point and line invariant transfer algorithms, dans *Proc. of the SPIE*, Vol. 3088, pp. 156–168.

BLANC, J. (1998), Synthèse de nouvelles vues d'une scène 3D a partir d'images existantes, Thèse de doctorat, Département Informatique, Institut National Polytechnique de Grenoble.

BOLLES, R. C., BAKER, H. H. et MARIMONT, D. H. (1987), Epipolar-plane image analysis: An approach to determoining structure from motion, *International Journal of Computer Vision* **1**(1), 7–55.

BURT, P. J. et ADELSON, E. H. (1983), A multiresolution spline with applications to image mosaics, *ACM trans. on graphics* **2**(4), 217–236.

CHEN, Q. et MEDIONI, G. (1997), Image synthesis from a sparse set of views, dans *Proc. Visualization'97*, pp. 269–275.

CHEN, S. E. (1995), QuickTime VR – an image-based approach to virtual environment navigation, dans *SIGGRAPH'95*, pp. 29–38.

CHEN, S. E. et WILLIAMS, L. (1993), View interpolation for image synthesis, dans *Computer Graphics Proceedings*, Anaheim, CA, USA, pp. 279–288.

CHEVRIER, C. (1997), A view interpolation technique taking into account diffuse and specular inter-reflections, *Visual Computer* **13**(7), 330–341.

COOKE, E., KAUFF, P. et SCHREER, O. (2002), Image-based rendering for teleconference systems, dans *WSCG02*, p. 119.

DYER, C. R. (1997), Image-based scene rendering and manipulation research at the university of wisconsin, dans *Proc. Image Understanding Workshop*, pp. 63–67.

EDELMAN, S. et BULTHOFF, H. H. (1992), Modeling human visual object recognition, dans *IJCNN International Joint Conference on Neural Networks*, Vol. 4, Baltimore, MD, USA, pp. 37–42.

FAUGERAS, O. et PAPADOPOULO, T. (1997), A nonlinear method for estimating the projective geometry of three views, rapport technique 3321, INRIA.

FUJIMURA, K. et MAKAROV, M. (1998), Foldover-free image warping, *Graphical Modeling and Image Processing* **60**(2), 100–111.

HARTLEY, R. I. (1992), Estimation of relative camera positions for uncalibrated cameras, dans *ECCV 1992*, pp. 579–587.

HARTLEY, R. I. (1997), In defense of the eight-point algorithm, *IEEE Transaction on pattern analysis and machine intelligence* **19**(6), 580–593.

HAVALDAR, P., LEE, M.-S. et MEDIONI, G. (1996), View synthesis from unregistered 2-d images, dans W. A. Davis et R. Bartels, éditeurs, *Proceedings. Graphics Interface '96*, Toronto, Ont., Canada, pp. 61–9.

HAVALDAR, P., LEE, M.-S. et MEDIONI, G. (1997), Synthesizing novel views from unregistered 2-d images, *Computer Graphics Forum* **16**(1), 65–73.

HEUNG-YEOP, J., JIN-HO, A., JE-HO, L., YONG-MOO, K., SANGKUK, K. et SANG-HUI, P. (1997), Fast interpolation technique on epipolar plane image using phase correlation, dans *Proceedings of 1997 IEEE International Symposius on Circuits ans Systems. Circuits and Systems in the Information Age. ISCAS '97*, Hong Kong, pp. 1425–1428.

HSU, R., KODAMA et K., HARASHIMA, H. (1994), View interpolation using epipolar plane images, dans *Proceedings ICIP-94*, Austin, TX, USA, pp. 745–749.

IRANI, M., ANANDAN, P. et HSU, S. (1995), Mosaic based representations of video sequences and their applications, dans *Fifth international conference on computer vision*, pp. 605–611.

KAMEI, K., MARUYAMA, M. et SEO, K. (1997), Scene synthesis by assembling striped areas of source images, dans *International Conference on Image Processing*, Vol. 2, pp. 482–485.

KANADE, T. et OKUTOMI, M. (1992), A locally adaptive window for signal matching, *International Journal of Compute Vision* **7**(2), 143–162.

KANG, S. (1997), A survey of image-based rendering techniques, rapport technique CRL 97/4, Digital Equipment Corporation, Cambridge Research Lab.

KUMAR, R., ANANDAN, P., IRANI, M., BERGEN, J. et HANNA, K. (1995), Representation of scenes from collections of images, dans *Proceedings IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA.

LAVEAU, S. et FAUGERAS, O. (1994), 3-D scene representation as a collection of images and fundamental matrices, rapport technique 2205, INRIA.

LHUILLIER, M. (1999), Towards automatic interpolation for real and distant image pairs, rapport technique 3619, INRIA Rhône-Alpes.

MANNING, R. A. et DYER, C. R. (1998), Dynamic view morphing, rapport technique 1387, University of Wisconsin, Madison, Wisconsin 53706.

MANSOURI, A. R. et KONRAD, J. (1997), Block-based winner-takes-all reconstruction of intermediate stereoscopic images, dans *Proc. of the SPIE*, Vol. 3309, pp. 922–933.

MCMILLAN, L. J. (1997), An Image-Based Approach to Three-Dimensional Computer Graphics, Thèse de doctorat, Computer Science Dept, University of North Carolina, Chapel Hill.

MILGRAM, D. L. (1975), Computer methods for creating photomosaics, *IEEE trans. on comp.* **C**(24), 1113–1119.

MILGRAM, D. L. (1977), Adaptive techniques for photomosaicking, *IEEE trans. on comp.* **C**(26), 1175–1180.

MOFFITT, F. H. et MIKHAIL, E. M. (1980), *Photogrammetry*, Harper & Row.

OWEN, C. B. et MAKEDON, F. (1997), Bottleneck-free separable affine image warping, dans *International Conference on Image Processing*, pp. 683–686.

PARK, J. I., YAGI, N. et ENAMI, K. (1994), Image synthesis based on estimation of camera parameters from image sequence, *IEICE Transactions on Information and Systems* **E77-D**(9), 973–986.

PELEG, S. (1981), Elimination of seams from photomosaics, *Comp. graphics and im. process.* **C**(16), 90–94.

PERRIER, J. S., AGAM, G. et COHEN, P. (2000*a*), Image-based view synthesis for enhanced perception in teleoperation, dans J. G. Verly, éditeur, *Proc. Enhanced and Synthetic Vision*, Vol. 4023 of *SPIE*, Orlando, Florida, pp. 213–224.

PERRIER, J. S., AGAM, G. et COHEN, P. (2000*b*), Physically valid triangulation of sparsely-matched images using texture information: application to view-synthesis, dans *Proc. Vision Interface*, Montreal, Canada, pp. 233–240.

QIAN CHEN, G. M. (1997), Image synthesis from a sparse set of views, dans *Visualization97*, pp. 269–275.

SCHARSTEIN, D. (1996), Stereo vision for view synthesis, dans *Proceedings 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 852–858.

SEITZ, S. M. (1997), Image-based transformation of viewpoint and scene appearance, Thèse de doctorat, Computer Science Dept, University of Wisconsin.

SEITZ, S. M. et DYER, C. R. (1995*a*), Complete sccene structure form four point correspondences, dans *Proc. Fifth Intl. Conf. on Computer Vision*, Cambridge MA, pp. 330–337.

SEITZ, S. M. et DYER, C. R. (1995*b*), Physically-valid view synthesis by image interpolation, dans *Proceedings IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA, pp. 18–25.

SEITZ, S. M. et DYER, C. R. (1996*a*), Toward image-based scene representation using view morphing, dans *Proceedings of the 13th International Conference on Pattern Recognition*, Vienna, Austria, pp. 84–89.

SEITZ, S. M. et DYER, C. R. (1996*b*), View morphing, dans *Computer Graphics Proceedings. SIGGRAPH '96*, New Orleans, LA, USA, pp. 21–30.

SEITZ, S. M. et DYER, C. R. (1997), View morphing: uniquely predicting scene appearance from basis images, dans *Proc. Image Understanding Workshop*, pp. 881–887.

SHASHUA, A. (1995), Algebraic functions for recognition, *IEEE trans. patt. anal. mach. intell.* **17**(8), 779–789.

SHASHUA, A. (1997), Trilinear tensor: the fundamental construct of multiple-view geometry and its applications, dans G. Sommer et J. J. Koenderink, éditeurs, *Algebraic Frames for the Perception-Action Cycle. International Workshop, AFPAC'97. Proceedings*, Kiel, Germany, pp. 190–206.

SHUM, H.-Y. et SZELISKI, R. (1997), Panoramic image mosaics, rapport technique MSR-TR-97-23, Microsoft Research.

SZELISKI, R. (1996), Video mosaics for virutal environments, *IEEE Comp. graphics and applications* pp. 22–30.

SZELISKI, R. et KANG, S. B. (1995), Direct methods for visual scene reconstruction, dans *IEEE Workshop on representations of visual scenes*, pp. 26–33.

SZELISKI, R. et SHUM, H.-Y. (1997), Creating full view panoramic image mosaics and environment maps, dans *Computer Graphics Proceedings, SIGGRAPH 97*, ACM, Los Angeles, CA, USA, pp. 251–8.

ULLMAN, S. et BASRI, R. (1991), Recognition by linear combinations of models, *IEEE Trans. PAMI* **13**(10), 992–1006.

WERNER, T., HERSCH, R. D. et HLAVAC, V. (1995), Rendering real-world objects using view interpolation, dans *Proceedings. Fifth International Conference on Computer Vision*, Cambridge, MA, USA, pp. 957–62.

WORLBERG, G. (1990), *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, California.

YONG, C., XUEHUI, L. et ENHUA, W. (1997), Image-based rendering: a technology for virtual reality system, *Journal of Software* **8**(10), 721–8.

ZHANG, Z. (1996), Determining the epipolar geometry and its uncertainty: A review, rapport technique 2927, INRIA.

ZHANG, Z., DERICHE, R., FAUGERAS, O. et LUONG, Q.-T. (1994), A robust technique for matching two uncalibrated images throught the recovery of the unkown epipolar geometry, rapport technique, INRIA.

ZHENG, Z., WANG, H. et TEOH, E. K. (1999), Analysis of gray level corner detection, *Pattern Recognition Letters* **20**(1), 149–162.